

Photonic computing enabled end-to-end image compression and reconstruction with adjustable ratio and high-fidelity

Corresponding Author: Professor Ang Li

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

The manuscript integrates a programmable 32×32 silicon photonic core with an opto-electronic pipeline and a lightweight reconstruction network (LiPICO-Net) to realize image compression and reconstruction with adjustable ratios. While the hardware effort is non-trivial, the present version falls short in methodological novelty and in fair/rigorous evaluation. Core claims rely on a trainable (deterministic) encoder rather than the random measurements customary to compressed sensing, making comparisons to prior CS-based networks not like-for-like. In addition, the end-to-end system is not truly realized experimentally, and the power/latency accounting appears optimistic by excluding critical mixed-signal components and I/O. I therefore cannot recommend publication in Nature Communications.

1. The lightweight performance of LiPICO-Net appears largely enabled by the jointly trained compression matrix (deterministic encoder), whereas classical CS baselines (ReconNet/DR2-Net, etc.) assume random measurements. Unless the authors hold the encoding scheme constant across methods, or compare against end-to-end learned compression (trainable encoder-decoder) baselines, the reported gains are not directly comparable and the conceptual advance over the literature remains limited.
2. The paper states that when the compression matrix changes, only the fully connected layer needs retraining. This still implies matrix-dependent fine-tuning, which weakens the claim of robustness and further undermines comparisons to methods designed to generalize across sampling matrices. Evidence of generalization to unseen matrices/scenes without any fine-tuning is needed.
3. To contextualize utility, please compare against JPEG/JPEG2000 and state-of-the-art learned compression codecs (e.g., Ballé/Cheng families). Report rate-distortion curves (bpp vs. PSNR/MS-SSIM) rather than isolated PSNR at a nominal “compression ratio”.
4. The system-level energy excludes/underplays GHz-class ADC/DAC/TIA power, which is often watt-level per channel at the cited 30–40 GHz bandwidths. Without a transparent component-level power table (modulators, heaters, PDs, TIAs, ADC/DAC, FPGA, drivers/clocking, thermal stabilization), the “2-3 orders of magnitude improvement” is not credible.
5. “49.5 ps/pixel” adds electronic and photonic compute times but ignores data motion and serialization/deserialization across ADC/DAC/FPGA and PCIe-like links. Given that the electronic path is actually on FPGA, while some profiling uses GPU, the estimate should be rebuilt with a consistent hardware model, including I/O and thermal/settling times for reprogramming the mesh.
6. The reconstruction is shown optically, while the compression stage remains electronic (as shown in Fig. 3), hence the claimed end-to-end photonic system is not demonstrated. A complete physical loop from optical encoding to reconstruction with real optics at both stages would materially strengthen the work.
7. There is a discrepancy in a cited latency figure for Ref. [1] (authors give 3.91 ps/pixel, whereas the source implies ≈ 0.98 ps/pixel under stated assumptions). All third-party metrics must be recalculated faithfully to avoid misleading comparisons.
8. A 32×32 optical core is small relative to megapixel imagery, please clarify tiling/streaming strategy, line-rate, buffer sizes,

and throughput limits when scaling beyond toy patches.

9. Report MS-SSIM/LPIPS at high ratios e.g., 256× where PSNR $\approx$ 17 dB is borderline useful. Provide visual error maps and task-level utility, such as detection/segmentation after reconstruction.

10. When reporting MAC/pixel reductions, ensure a hardware-consistent mapping, e.g., analog MAC equivalence, bit-precision, quantization, overhead of calibration and readout.

11. Numerous typos/grammar issues such as “computtaional”, “haven’t to retrain total parameters”, unit typos “10.58 pJ/pixel” vs “nJ/pixel”, should be corrected, figures need axes/units and consistent legends.

In summary, the manuscript demonstrates engineering effort but, as written, lacks methodological novelty, fair evaluation, and credible system-level accounting commensurate with Nature Communications. A substantially revised version with like-for-like baselines, full system demonstration, rigorous power/latency tables, and rate-distortion evaluation might be suitable for a more specialized venue.

Reference

[1] X. Wang, B. Redding, N. Karl, C. Long, Z. Zhu, J. Skowronek, S. Pang, D. Brady, and R. Sarma, "Integrated photonic encoder for low power and high-speed image processing," Nat. Commun. 15, 4510 (2024).

Reviewer #2

(Remarks to the Author)

The authors demonstrate a programmable 32 by 32 linear operator for optical signal transformation. They demonstrate that the latency per input signal is less than 50 ps and an encoding power per pixel of 10 pJ. High speed tensor operations implemented by this chip could be used for many applications, here the authors choose to focus on image data compression and decompression. This is a powerful application for this device, it is particularly interesting that the authors here demonstrate the ability to load reconstruction weights on to the device. End to end, this enables a significant energy per pixel reduction relative to the passive encoding strategy implemented in their reference 28. As noted in the manuscript, flexible and adaptive compression ratios are useful in these applications. While one could consider numerous variations on how this photonic tensor operator might be applied, the approach demonstrated here is sound, the paper is well written and the results are interesting. This work is of particular interest because it is an end-to-end demonstration of very practical application of photonic neural processing.

Reviewer #3

(Remarks to the Author)

This work presents an image compression and reconstruction scheme based on a programmable photonic processor. The fundamental part of the work is very similar to the previously published study (X. Wang et al., “Integrated photonic encoder for low-power and high-speed image processing,” Nature Communications 15, 4510 (2024)). As a novelty, the authors provide two new contributions:

- First, the authors propose and experimentally demonstrate the use of a 32×32 programmable mesh based on Mach–Zehnder interferometers with an FFT architecture to implement random matrix multiplication, which presents advantages in terms of reconfigurability when compared to random scattering devices such as those previously demonstrated.
- Second, the authors use an alternative neural network algorithm for the reconstruction of the encoded images, which provides comparable results to more complex state-of-the-art solutions.

While the work is of interest to the community and is clearly written, I cannot recommend the publication of the article in its current form, as the manuscript omits some important points and includes some misleading statements, especially those related to the photonic processor and its capabilities:

- Throughout the article, the authors claim several times that this is the first demonstration of a compression and reconstruction approach using an optoelectronic computing processor. However, as they themselves mention, a similar work was previously published, which is also optoelectronic. The authors should revise this statement.
- Regarding the novelty of the work, the authors focus their comparison on the differences with X. Wang et al., “Integrated photonic encoder for low-power and high-speed image processing,” Nature Communications 15, 4510 (2024). In the photonics domain, the fundamental difference lies in the programmability of the photonic processor. The authors propose the use of an FFT-based architecture, which is less complex than other solutions such as the rectangular (Clements) or triangular (Reck) configurations, at the expense of reduced matrix accessibility. I wonder whether this limitation in the number of matrices that can be implemented may represent a fundamental restriction for achieving better results. A comparison with other photonic architectures should be included in this regard. Moreover, the authors should provide more information about the architecture used, as well as its benefits and limitations.

- Regarding the photonic processor:

- o The authors mention the inclusion of 32 high-speed modulators and photodetectors that allow a working speed of more than 30 GHz. They also provide the experimental response of these devices. During the experiments, which signal frequency was used? In case lower-frequency signals were employed, the authors should adapt their metrics (energy/pixel, latency, etc.) and present both the experimental and the achievable figures.
- o Little information is provided regarding the photonic building blocks, which reduces the reproducibility of the work. System

insertion losses (MZIs, optical packaging, etc.) should be reported.

o The authors state that the photonic chip is co-packaged with the necessary control electronics, but I do not see the electronic elements in Fig. a–b. This statement, which is presented as a novelty, should be revised if the control electronics are independent from the optical processor.

o A 32×32 optical processor is at the edge of state-of-the-art photonic systems; however, it would be beneficial to include a scalability study of the photonic processor.

- Regarding LiPICO-Net, as I understand, it was adapted to work with the 32×32 size of the processor. How does it scale with the size of the input image?

- Finally, concerning the estimation of performance metrics in the Supplementary Material, the authors mention that the energy contribution of the electronics such as ADC/DAC/TIA is negligible. However, the electrical part is usually the most power-consuming in photonic processors (<https://pubs.aip.org/aip/app/article/7/2/020902/2835121/Scaling-up-silicon-photonic-based-accelerators>, <https://www.nature.com/articles/s41586-025-08854-x>). The authors should include the impact of the electronic part and adapt their comparisons accordingly.

- Minor: There is a typo in the abstract where the authors write “computational.”

Reviewer #4

(Remarks to the Author)

The manuscript presents a significant advancement in photonic computing applied to image processing. The Authors demonstrate, for the first time, an end-to-end image compression and reconstruction system based on a fully integrated opto-electronic computing processor (OECF), achieving latency and energy efficiency improvements of several orders of magnitude compared with conventional GPU-based methods. The work is ambitious, timely, and highly relevant to the fields of silicon photonics, optical computing, and computational imaging. The key novelty lies in the co-integration of a 32×32 programmable photonic matrix core with high-speed modulators, detectors, and control electronics, enabling both compression and reconstruction in a unified photonic–electronic pipeline. The development of the lightweight LiPICO-Net, specifically optimized for photonic hardware constraints, further enhances the originality. Compared to prior work such as Wang et al. [28], this study achieves adjustable compression ratios and full integration, which constitute major improvements.

Here, my major comments:

1. The methodology is carefully described in both the main text and supplementary materials. The description of LiPICO-Net (Supplementary S1) is comprehensive, including the convolutional architecture and mathematical formulation, while the details of the opto-electronic hardware (Supplementary S2) and performance estimations (S3) are clearly presented. However, several methodological aspects would benefit from further elaboration to reinforce the technical soundness of the work.

a. The reported latency and energy consumption values rely on idealized assumptions regarding the modulation bandwidth, optical power levels, and overall system clock rate. It is not entirely clear how these parameters were experimentally verified or constrained by measurement. For instance, while the modulator and detector bandwidths are reported as exceeding 30 and 40 GHz respectively, it would be valuable to discuss whether the 30 GHz operational frequency (used to estimate the MAC time of 0.032 ps) represents a measured limit or a theoretical upper bound. A brief uncertainty analysis or sensitivity assessment—quantifying how variations in optical power or bandwidth impact latency and energy per pixel—would significantly strengthen the reliability of the reported performance metrics.

b. The validation of the reconstruction quality relies exclusively on the Peak Signal-to-Noise Ratio (PSNR). While PSNR remains a standard metric, it often fails to capture perceptual fidelity or structural accuracy in complex images. Including additional metrics such as Structural Similarity Index (SSIM) or Learned Perceptual Image Patch Similarity (LPIPS) would provide a more robust and perceptually relevant evaluation of the reconstruction quality, particularly at high compression ratios where visual artifacts may dominate.

c. The reported total energy consumption per pixel appears to primarily account for the photonic matrix operations. However, the OECF includes multiple electronic components—FPGA, ADCs, DACs, and TIAs—that can contribute significantly to total energy consumption, especially in high-speed operation. It should be clarified whether these electronic contributions were measured, estimated, or neglected in the presented calculations. If omitted, the authors should specify this explicitly and discuss how their inclusion would affect the overall system-level efficiency.

2. The manuscript frequently refers to “computing” and “computation” in the context of photonic matrix multiplications. While this is technically correct, it would be important to stress that the scope of photonic computing extends well beyond linear algebraic operations. In the broader context of optical information processing, “computing” encompasses a wide range of primitives, including discrete and continuous Fourier transforms (FFT), convolutions, correlations, and nonlinear mappings, which are increasingly realized through integrated photonic architectures (see, for example, doi: 10.1109/JLT.2024.3453670). By acknowledging this broader framework, the authors could more clearly position their work within the expanding field of hybrid analog–digital photonic computing, highlighting how the proposed OECF architecture might be adapted to more general-purpose or signal-processing-oriented tasks in the future.

3. The performance results are impressive and generally well supported by quantitative data. The comparison in Table 1 effectively summarizes the advantages over GPU-based models, both in latency and energy efficiency. Nevertheless, a deeper and more critical discussion of scalability and practical implementation limits would improve the manuscript.

a. Scalability represents a key challenge for photonic matrix processors. As the number of optical channels increases beyond 32×32, phase stability, thermal crosstalk, and calibration complexity can grow nonlinearly. The authors should discuss to what extent the FFT-based mesh architecture remains practical for larger matrices and whether emerging alternatives—such as modular photonic tensor cores or folded mesh layouts—might offer improved scalability.

b. The manuscript should clarify how system performance scales with compression ratio, particularly near the extreme

regime of $256\times$. At such high ratios, optical noise, quantization effects, and the conditioning of the compression matrix may significantly affect reconstruction quality. Quantitative results on PSNR degradation or convergence behavior during training across different ratios would provide valuable insight into system robustness.

c. The degradation observed between simulated (AE) and experimental (RM) results in Fig. 3(b) deserves more attention. It would be helpful to identify the dominant physical sources of this discrepancy—such as phase noise, coupling losses, MZI imbalances, or limited precision in the programmable phase shifters—and to discuss potential mitigation strategies, including calibration protocols or feedback control.

4. The figures are generally clear, well organized, and of high visual quality. Figure 4(g) provides a convincing illustration of the reconstruction performance at different compression ratios. Nonetheless, some improvements to data presentation would increase the transparency of the results.

a. Each figure caption should clearly distinguish between simulated data, idealized results, and experimentally measured quantities. For instance, specifying “simulation (AE)” or “experimental (RM)” in the captions themselves would help readers interpret the data correctly without needing to refer back to the text.

b. Axis labels and units in all quantitative plots, particularly those showing PSNR or latency versus compression ratio, should be explicitly stated to ensure reproducibility and interpretability.

5. The supplementary information (Sections S1–S3) is well structured and provides essential implementation details. The mathematical derivation of LiPICO-Net and the step-by-step energy and latency estimations are valuable. However, cross-referencing between the main manuscript and supplementary materials could be made more systematic. For example, when equations or design parameters are derived in the supplementary sections, the main text should explicitly refer to the corresponding section or equation number. This would enhance readability and help readers follow the logical connection between modeling assumptions and experimental implementation.

6. The manuscript is overall well written and logically structured. The exposition is clear, and the technical depth is appropriate for a high-impact journal. However, several stylistic and minor linguistic improvements are recommended. A few typographical errors persist (e.g., “computaional,” “lantency”), and long compound sentences could be simplified to improve clarity. The introduction would also benefit from a short paragraph connecting the presented work to recent advances in photonic tensor processors and AI accelerators, which would situate the contribution within the rapidly growing landscape of photonic–machine learning convergence.

Minor comments:

- Clarify whether the reported 10.58 nJ/pixel total energy includes both optical and electronic contributions or refers exclusively to the optical computation segment.
- Provide a brief discussion on the reconfiguration speed and programmability of the phase shifters, commenting on the trade-offs between thermo-optic tuning (high power, slow response) and potential alternatives such as carrier-injection or microheater-free architectures.
- Include uncertainty estimates or error bars for key quantitative parameters such as latency, energy per MAC, and PSNR to substantiate the numerical claims and convey experimental repeatability.
- Correct remaining typographical errors and ensure consistent notation throughout the manuscript (e.g., uniform use of subscripts in PSNR equations, correct formatting of units, consistent use of “ $\times$ ” versus “x” for dimensional notation).
- Consider adding a concise summary table in the Supplementary Information comparing the theoretical and experimental parameters of the OECP (modulator bandwidth, PS efficiency, photodetector responsivity, FPGA timing, etc.) for clarity and reproducibility.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

The authors' responses fall short of addressing the concerns raised. Many key issues were not addressed directly, and several critical questions remain unanswered. The rebuttal does not resolve the fundamental doubts regarding the validity and credibility of the work. Given the lack of clear novelty and the insufficient methodological rigor, this work is not suitable for publication in Nature Communications.

1. The proposed work essentially applies on-chip photonic interferometric computing to the task of image compression and reconstruction. The underlying methodology follows approaches that are already well established in the field. Moreover, the proposed system still relies on electronic computation to implement FC layers, residual connections, yet these electronic components are selectively excluded from the performance and energy evaluations, which is confusing and undermines the rigor of the analysis.

2. In the response to the comparisons with JPEG, the authors argue that the proposed work is fundamentally different from conventional compression methods. I respectfully disagree with this viewpoint. The authors are expected to comprehensively evaluate the proposed system. Moreover, the authors themselves cite figures from Ref. [1] demonstrating that photonic image compression can achieve performance comparable to JPEG in certain scenarios. This evidence, in fact, underscores the necessity of including comparisons with JPEG and other classical algorithms.

3. The authors argue that excluding AD/DA power from the energy analysis is common practice. However, even in the cited works, this exclusion is typically justified only when reporting core optical computation energy or when AD/DA occurs solely at the system output. In contrast, system-level energy evaluations generally include AD/DA and related electronic components. Ref. [2] reports modulator power consumption, Ref. [3] focuses on frequency comb power; Ref. [4] considers

convolution operations only and explicitly emphasizes “considering only optical contributions”, and Ref. [1] reports solely the power of the image compression module.

The manuscript claims system-level energy efficiency, yet the proposed system requires AD/DA conversions within the processing pipeline rather than only at the output. As such, the associated energy cannot be neglected. While the power of the compression and reconstruction modules in Table 1 may remain unchanged, the end-to-end system energy should include the overhead between these stages. The statement in the Supplementary Materials that “ADC/DAC/TIA contributes negligible energy consumption” is also potentially misleading and should be revised.

4. The latency estimation for other related works appears to adopt calculation methods that differ from those used in the original references. The authors may need to explicitly describe the adopted latency calculation methodology in the Supplementary Materials.

5. The system energy is reported as 10.58 nJ/pixel in the Introduction and energy analysis, but as 10.58 pJ/pixel in the Discussion. If this is not a typographical error, the authors should clarify the discrepancy and provide the relevant calculations.

Related original text: “Heterogeneous integration of all necessary control electronics establishes a unified photonic-electronic pipeline delivering unprecedented latency (49.5 ps/pixel) at 10.58 pJ/pixel.”

The performance metrics in the present form are evaluated in an unfair manner, and the claimed innovation does not demonstrate clear advantage over Ref. [1] and related work. The manuscript does not meet the publication standards of Nature Communications.

Reference

- [1] X. Wang, B. Redding, N. Karl, C. Long, Z. Zhu, J. Skowronek, S. Pang, D. Brady, and R. Sarma, "Integrated photonic encoder for low power and high-speed image processing," *Nat. Commun.* 15, 4510 (2024).
- [2] Y. Shen, N. C. Harris, S. Skirlo, M. Prabhu, T. Baehr-Jones, M. Hochberg, X. Sun, S. Zhao, H. Larochelle, D. Englund, and M. Soljačić, "Deep learning with coherent nanophotonic circuits," *Nat. Photonics* 11, 441–446 (2017).
- [3] X. Xu, M. Tan, B. Corcoran, J. Wu, A. Boes, T. G. Nguyen, S. T. Chu, B. E. Little, D. G. Hicks, R. Morandotti, A. Mitchell, and D. J. Moss, "11 TOPS photonic convolutional accelerator for optical neural networks," *Nature* 589, 44–51 (2021).
- [4] J. Feldmann, N. Youngblood, M. Karpov, H. Gehring, X. Li, M. Stappers, M. Le Gallo, X. Fu, A. Lukashchuk, A. S. Raja, J. Liu, C. D. Wright, A. Sebastian, T. J. Kippenberg, W. H. P. Pernice, and H. Bhaskaran, "Parallel convolutional processing using an integrated photonic tensor core," *Nature* 589, 52–58 (2021).

Reviewer #3

(Remarks to the Author)

The authors have addressed all my concerns. The manuscript is ready for publication in Nature Communications.

Reviewer #4

(Remarks to the Author)

After reviewing the full response document, the authors appear to have answered all reviewer questions in a complete and structured manner. For each reviewer, the authors directly addressed every numbered and sub-question, provided clarifications, added missing data where requested, and justified methodological choices. In several cases they expanded figures, added supplementary tables, corrected typographical issues, and integrated additional analyses such as PSNR/SSIM/MS-SSIM curves, scalability discussions, power-consumption clarifications, and system-level limitations.

Version 2:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

The response addresses my concern. I agree the publication of this manuscript in Nature Communications.

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

Reviewer #1 (Remarks to the Author):

The manuscript integrates a programmable 32×32 silicon photonic core with an opto-electronic pipeline and a lightweight reconstruction network (LiPICO-Net) to realize image compression and reconstruction with adjustable ratios. While the hardware effort is non-trivial, the present version falls short in methodological novelty and in fair/rigorous evaluation. Core claims rely on a trainable (deterministic) encoder rather than the random measurements customary to compressed sensing, making comparisons to prior CS-based networks not like-for-like. In addition, the end-to-end system is not truly realized experimentally, and the power/latency accounting appears optimistic by excluding critical mixed-signal components and I/O. I therefore cannot recommend publication in Nature Communications.

We appreciate Reviewer #1's evaluations of our manuscript. Even if we respectfully disagree with some critical comments, we treat every single comment seriously and provide as many details as possible in the response.

1. The lightweight performance of LiPICO-Net appears largely enabled by the jointly trained compression matrix (deterministic encoder), whereas classical CS baselines (ReconNet/DR2-Net, etc.) assume random measurements. Unless the authors hold the encoding scheme constant across methods, or compare against end-to-end learned compression (trainable encoder-decoder) baselines, the reported gains are not directly comparable and the conceptual advance over the literature remains limited.

We thank the reviewer for this insightful comment regarding the comparison methodology. We respectfully disagree with the view that the gains are not directly comparable or that the conceptual advance is limited, and we provide the following points for consideration:

1) We agree that the ideal CS theory often uses perfectly random matrices. However, the fundamental principle of CS is to exploit signal sparsity for sub-Nyquist sampling, rather than mandating perfect randomness as an absolute requirement. In practice, many successful CS implementations use structured or pseudo-random measurement matrices (e.g., partial Fourier, Bernoulli, or matrices from physical systems) that are sufficiently incoherent with the signal basis, without being perfectly random. Here are a few examples in CS-based spectral analysis, where pseudo-random matrix were used[1-3]. Our photonic chip generates a programmable, high-speed, and low-power measurement matrix that possesses the necessary stochastic properties (i.e., it is non-periodic with certain randomness) to function effectively as a CS encoder. This approach is not only valid but also represents a practical advancement, as generating ideal random matrices in hardware is challenging. **The ability of our system to perform high-fidelity reconstruction without relying on perfect randomness should be viewed as a strength, enhancing its practicality for real-world deployment.** In other

words, if a model can only work with perfect random matrix instead of matrix generated by physical system, I would rather claim its advancement to be limited.

2) While the jointly trained, photonic-generated matrix optimizes overall system performance, our reconstruction network **can also work effectively with ideal random matrices** (e.g., Gaussian matrices the same as used in ReconNet and DR2-Net). As shown in the simulations in Fig. 3(b) (LiPICO-Net (AE)), which used random Gaussian matrices for compression, LiPICO-Net achieves performance comparable to ReconNet and even surpasses DR2-Net at very low sampling rates. This demonstrates that the lightweight nature and efficacy of LiPICO-Net are not solely dependent on a custom-trained matrix but are inherent properties of the network architecture itself. We believe the reviewer might have missed this part.

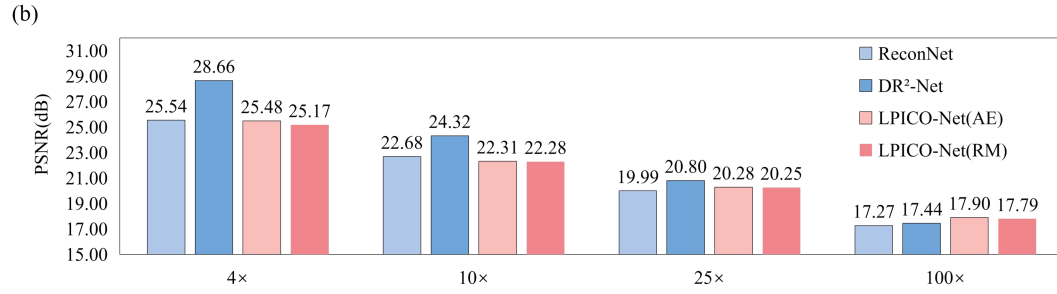


Fig. R1 Performance comparison between our LiPICO-Net with other seminal models including ReconNet and Dr2-Net at four compression ratios. AE stands for “as expected” which is obtained through simulations (random Gaussian matrix) and RM represents “real measurements”, which is obtained from experiments.

The original context in the manuscript is copied here:

“We evaluated LiPICO-Net's theoretical performance through PyTorch simulations using **random Gaussian compression matrices**. The dataset for training is the same 91 images dataset used in[29] and the dataset for testing is the 11 images dataset used in[30]. 4 different sampling rates (e.g. compression ratio) were studied: 25%, 10%, 4%, and 1%. And we provide direct performance comparison with two seminal AI models for image reconstruction: ReconNet and Dr2-Net. The mean PSNR results are shown in Fig. 3(b)LiPICO-Net(AE), where AE stands for “as expected”. Simulation results reveal LiPICO-Net matches ReconNet's PSNR across sampling rates and surpasses DR2-Net at 1% sampling. Due to discrepancy between the ideal matrix and the generated matrix of PCC, the mean PSNR based on physical implementation of LiPICO-Net on PCC (but image compression is still in electronics) decrease a little compared with the LiPICO-Net(AE), as shown in Fig. 3(b)LiPICO-Net(RM), where RM is short for “real measurement”, but still maintain at a high level.”

Once again, the feature that the compression matrix can be co-optimized in the entire network shall be considered as major advancement than models that lack of such ability and can only work with perfect random matrix.

3) From an end-user or system designer's standpoint, the ultimate performance metrics—such as reconstruction quality (PSNR), computational costs, latency, and energy consumption—are of paramount importance. Our comparison aims to provide a realistic benchmark between a state-of-the-art integrated photonic-electronic system (our work) and established software-based algorithms (ReconNet/DR2-Net) running on powerful GPUs. This comparison is highly relevant for applications where power, speed, and integration are critical constraints, as it demonstrates the tangible benefits of our co-designed hardware-software approach, regardless of the specific encoding matrix used.

In summary, we believe our work provides a significant conceptual and practical advance by demonstrating a fully integrated opto-electronic system that co-designs a programmable physical encoder (the PCC) with a hardware-optimized reconstruction network (LiPICO-Net). This co-design is key to achieving the unprecedented end-to-end latency and energy efficiency reported, which we believe is a transformative step for computational imaging systems.

Ref:

1. Nature Photonics volume 18, pages1312–1319 (2024)
2. Optics Letters Vol. 42, Issue 1, pp. 25-28 (2017)
3. Nature Communications volume 12, Article number: 2704 (2021)

2. The paper states that when the compression matrix changes, only the fully connected layer needs retraining. This still implies matrix-dependent fine-tuning, which weakens the claim of robustness and further undermines comparisons to methods designed to generalize across sampling matrices. Evidence of generalization to unseen matrices/scenes without any fine-tuning is needed.

We appreciate the reviewer's comment regarding model generalization and fine-tuning. We would like to clarify the practical considerations and the robustness of our approach:

1) The requirement to adapt part of the network, specifically the initial fully-connected layer (FCL), when the compression ratio or matrix changes is a standard and well-established practice in the field of deep learning-based compressive sensing, even. This is fundamentally due to the change in the input dimension of the reconstruction network when the number of measurements varies. Seminal works like ReconNet employ precisely the same strategy: they keep the convolutional layers fixed and retrain only the first FCL when the sampling rate is altered. This is not a limitation unique to our model but a structural necessity for handling variable compression ratios. Therefore, our approach is directly comparable to these established baselines on an equal footing. Besides, training FCL requires trivial time as discussed in [1]:

Specifically, we adapt the convolutional layers (C1-C6) of a pre-trained network for the same or slightly higher MR, hence forth referred to as the base network, and train only the fully connected (FC) layer with random initialization for 1000 iterations (or equivalent time of around 2 seconds on a Titan X GPU), while keeping C1-C6 fixed. The mean PSNR(without BM3D) for the test-set at various MRs, the time taken to train models and the MR of the base network are given in table 3. From the table, it is clear that the overhead in computation for new Φ is trivial, while the mean PSNR values are comparable to the ones presented in table 1.

2) In practical applications, it is highly unrealistic to expect any high-performance AI model to operate across vastly different sensing configurations (e.g., different compression ratios or matrices generated by different hardware instances) without any form of adaptation or calibration. The key metric for robustness in this context is the efficiency and extent of this required adaptation. Our model demonstrates strong robustness by requiring only the FCL to be retrained, which is a rapid and computationally inexpensive process compared to retraining the entire network (including the convolutional layers executed on the photonic core). This represents a significant advantage in resource-constrained edge scenarios, where storing multiple full models for different settings is impractical, but a quick fine-tuning of a single layer is feasible.

Ref:

1. K. Kulkarni, et al., "Reconnet: Non-iterative reconstruction of images from compressively sensed measurements," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016), 449-458.

3. To contextualize utility, please compare against JPEG/JPEG2000 and state-of-the-art learned compression codecs (e.g., Ballé/Cheng families). Report rate-distortion curves (bpp vs. PSNR/MS-SSIM) rather than isolated PSNR at a nominal "compression ratio".

We thank the reviewer for this suggestion, which helps to better situate our work within the broader field of image compression. We fully adopt the suggestion to report rate-distortion curves (bpp vs. PSNR/MS-SSIM) rather than isolated PSNR at a nominal "compression ratio" but we would like to respectfully clarify why a direct comparison with JPEG may not be applicable in this context.

- 1) Regarding additional results on rate-distortion curves of our approach: Firstly we would like to clarify a practical difficulty in directly obtaining the rate-distortion curves on our current photonic system: while our compression ratio can be tuned by controlling the sum of output ports on the chip, it is challenging to achieve arbitrary ratios (e.g., exactly $7\times$ or $11\times$) due to the analog nature of the computation. This makes it difficult to accurately sample the continuous rate-distortion function required for a conventional bpp curve. To address this and still provide a comprehensive performance evaluation, we utilize LiPICO-Net(AE) which can easily achieve accurate compression ratios for verification and have accordingly expanded our

tests with more compression ratios to generate smoother curves. We believe these results offer a robust and thorough evaluation of our method's capabilities. The results are as followings and will be included in the Supplementary Information:

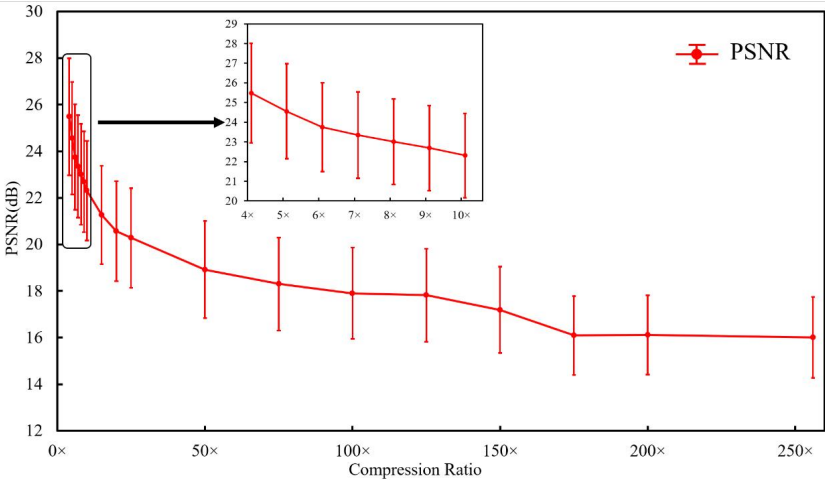


Fig. R1 Average PSNR versus Compression Ratio of 11 images.

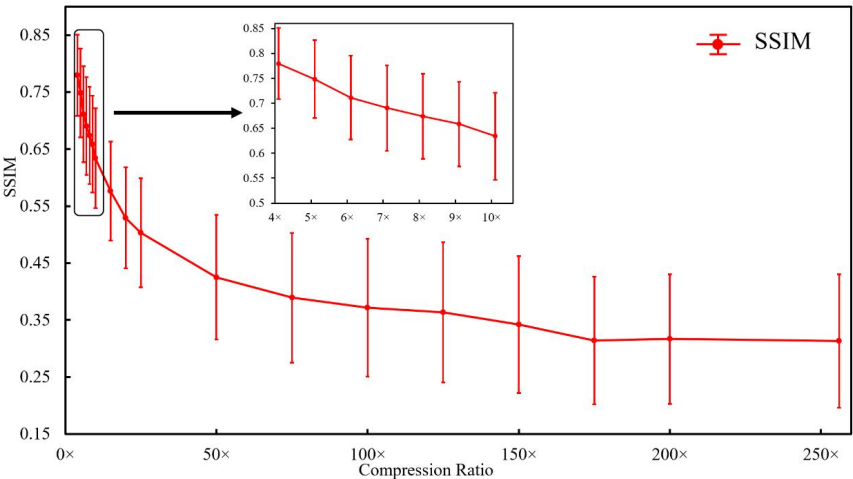


Fig. R2 Average SSIM versus Compression Ratio of 11 images.

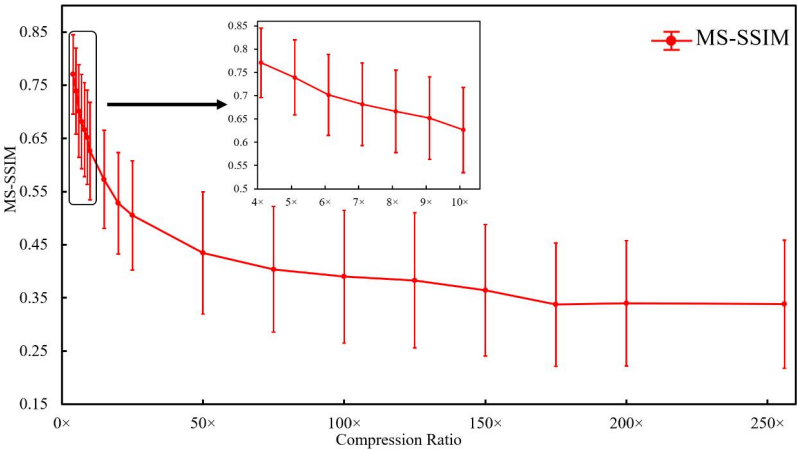


Fig. R3 Average MS-SSIM versus Compression Ratio of 11 images.

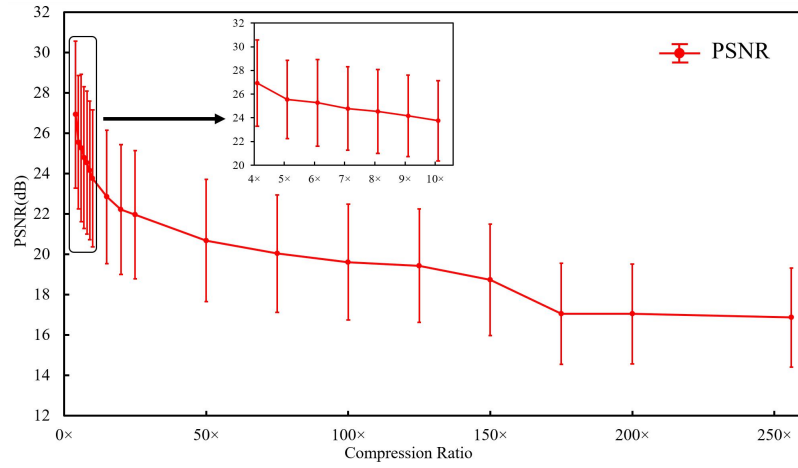


Fig. R4 Average PSNR versus Compression Ratio of BSDS 200.

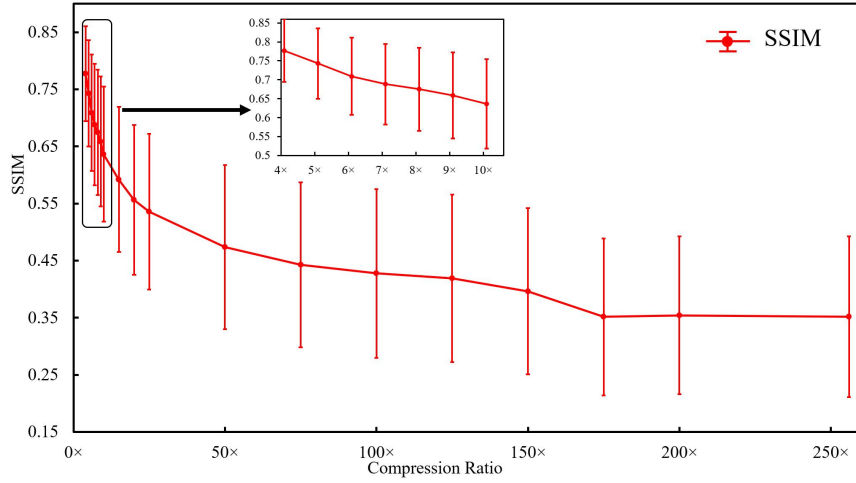


Fig. R5 Average SSIM versus Compression Ratio of BSDS 200.

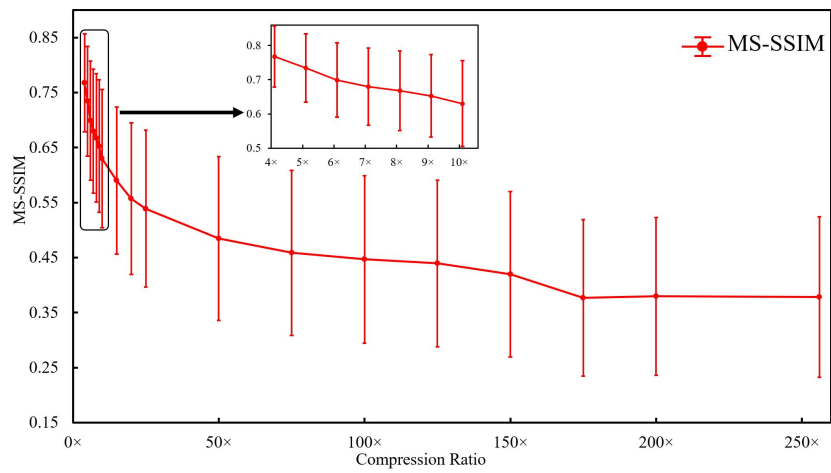


Fig. R6 Average MS-SSIM versus Compression Ratio of BSDS 200.

2) Regarding the comparison with JPEG/JPEG2000: Our work demonstrates a **computational** imaging and compression system based on a photonic processor. The core innovation lies in using optics to perform the compressive sensing (the matrix multiplication) at ultra-high speed and low power, followed by a lightweight AI-based reconstruction. This is fundamentally different from the **transform-and-entropy-coding paradigm of JPEG/JPEG2000 and the end-to-end learned compression codecs** (e.g., Ballé et al.), which are purely digital algorithms designed to minimize bitrate for a given distortion. We will thoroughly clarify why we believe that comparing with algorithms such as JPEG/JPEG2000 may not directly align with the goal of highlighting our specific contributions.

- Our work builds directly upon the field of deep learning-based compressive sensing (CS), pioneered by works such as ReconNet and DR2-Net. It is standard practice in this sub-field to report performance against other CS reconstruction algorithms (as we have done) using PSNR vs. measurement rate (and included rate-distortion curves), as the primary goal is to demonstrate improved reconstruction quality from a compressed image. Comparisons against JPEG are less common in this specific context because the paradigms are so distinct.
- Classical codecs operate on and produce bits. Our system, in its presented form, operates on and produces analog measurements (before ADC). The "compression ratio" in our work refers to the reduction in the number of these analog measurements, not the final bitrate. A direct R-D comparison (bpp vs. PSNR) would require **adding a subsequent quantization and entropy coding stage to our system**, which is a significant separate research topic and was not the focus of this work, which aimed to demonstrate the photonic computational core.
- The primary metrics for our system are throughput, latency, and energy consumption for the computational compression and reconstruction process. We target scenarios where these metrics are paramount, such as real-time processing of massive remote sensing images where electronic systems "falter due to power and latency constraints" (as stated in our Abstract). In these contexts, the computational cost of running a complex codec like JPEG2000 or a Ballé-style model on a digital processor can be the limiting factor, even if their compression efficiency is high.
- The working scenarios or performance limit of our approach and JPEG are very distinct. JPEG aims to provide more accurate image reconstruction at a lower compression ratio and processing speed. While our approach aims to provide real-time processing of large image with much higher compression ratio and lower latency. As shown in [1], where they already provide performance comparison between photonic-enabled image compression vs JPEG, the max. compression ratio of JPEG is around 50, while our approach can support over 200x ratio.

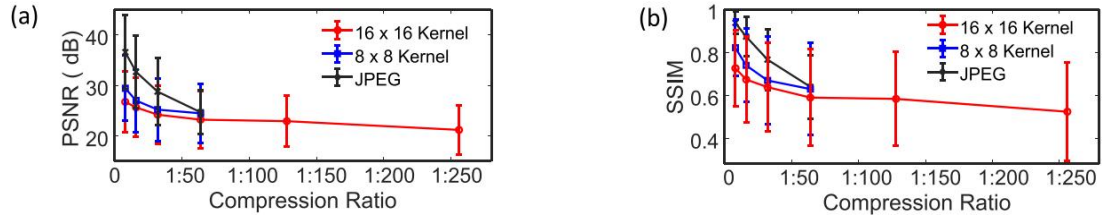


Fig. R7 Quote from [1]: Comparison of photonic image compression to digital JPEG image compression using images from the DIV2K and Flickr2K dataset^{31,32}. a Comparison of mean PSNR of the reconstructed images from the test dataset (sample size = 500 test images) and images compressed using the JPEG algorithm as a function of compression ratio. The error bars correspond to 1 standard deviation. The blue (red) line corresponds to photonic compression with an 8×8 (16×16) kernel. The black line corresponds to JPEG compression. The data for compression ratio of 1: 64 for JPEG compression has a smaller sample size of only 400 images as ~ 100 test images could not be compressed beyond 1:45 using the JPEG algorithm. b Comparison of mean SSIM of the reconstructed images from the test dataset (sample size = 500 test images) and images compressed using the JPEG algorithm as a function of compression ratio. The error bars correspond to 1 standard deviation. The blue (red) line corresponds to photonic compression with an 8×8 (16×16) kernel. The black line corresponds to JPEG compression. The data for compression ratio of 1: 64 for JPEG compression has a smaller sample size of only 400 images as ~ 100 test images could not be compressed beyond 1:45 using the JPEG algorithm.

Ref:

1. Nature Communications volume 15, Article number: 4510 (2024)

4. The system-level energy excludes/underplays GHz-class ADC/DAC/TIA power, which is often watt-level per channel at the cited 30–40 GHz bandwidths. Without a transparent component-level power table (modulators, heaters, PDs, TIAs, ADC/DAC, FPGA, drivers/clocking, thermal stabilization), the “2-3 orders of magnitude improvement” is not credible.

We thank the reviewer for raising this critical point regarding energy accounting. We acknowledge that a comprehensive, component-level power breakdown is ideal for a full system-level comparison. However, we would like to clarify the scope of our claim and the standard practices in the emerging field of photonic computing.

1) The primary contribution of our work is the demonstration of a monolithically integrated photonic core (PCC) that performs the computationally intensive matrix multiplication at the speed of light with ultra-low latency and power. The standard and most critical metric for evaluating such a photonic computing core is its intrinsic performance – the energy and latency for the fundamental optical operation (matrix-vector multiplication), as this directly showcases the photonic advantage. This approach of reporting the core photonic processor's

performance is well-established in the field [e.g., Refs. 17, 21, 22, 27, 28 in our manuscript], as it allows for a direct comparison of the photonic technology itself, independent of the peripheral electronics which can be optimized separately and follow their own Moore's Law scaling. Here are some concrete examples of performance characterization for PCC:

Shen et al. [1] calculate only the energy consumption of resistances on phase modulators:

Once all parameters have been trained and programmed on the nanophotonic processor, forward-propagation computing is performed optically on a passive system. In our implementation, maintaining the phase modulator settings requires some (small) power of ~ 10 mW per modulator, on average.

Xu et al. [2] only take the energy consumption of frequency comb they used into consideration:

The soliton crystal microcomb provided more than 90 channels over around 36 nm of the telecommunications C-band (1,540–1,570 nm), offering adiabatically generated low-noise frequency comb lines with a small footprint of <1 mm² and potentially low power consumption (<100 mW using the technique in ref. 31).

Feldmann et al. [3] consider only optical energy contributions:

A key feature of our approach is that, because the convolutional operation is a passive transmission measurement, the calculations can in theory be performed at the speed of light at very low power (17 fJ per MAC, considering only optical contributions)

Similarly, References [4-6] also accounted for only the optical energy consumption.

2) The pioneering work in [7], which is the first study to use photonic chip for image compression, doesn't incorporate the contribution from necessary electronics/computers when claiming "exceeding Terapixel/second using less than 100 fJ/pixel" and "this technique has the potential to encode Terapixel/second data streams utilizing <100 femtojoules per pixel—representing a $>1000\times$ reduction in power consumption compared with the state-of-the-art electronic approaches", either. Since we establish our work on the foundation of this pioneer work, we believe it's better to use the same standard.

3) The reviewer correctly points out the power challenges of high-speed electronics. We emphasize that the 30-40 GHz bandwidths measured for the modulators and detectors (Figs. 4d,e) represent the component-level speed capability, demonstrating the platform's potential for extreme throughput. The actual operating speed and thus the power of the ADCs/DACs in our current proof-of-concept demonstration are dictated by the FPGA and the image processing pipeline, not the ultimate photonic bandwidth. Future iterations of this work, aimed at maximizing throughput, would indeed require co-design with advanced ASICs for the electronics. Nevertheless, even with the current setup, the system-level energy efficiency is transformative for the targeted applications.

But we have included additional discussions in the Supplementary Information for address the reviewer's concern.

Ref:

1. Nature photonics 11, 441-446 (2017).
2. Nature 589, 44–51 (2021).
3. Nature 589, 52–58 (2021).
4. Nat Commun 14, 70 (2023).
5. Nature 569, 208-214 (2019).
6. Nature communications 13, 2131 (2022).
7. Nature Communications volume 15, Article number: 4510 (2024)

5. "49.5 ps/pixel" adds electronic and photonic compute times but ignores data motion and serialization/deserialization across ADC/DAC/FPGA and PCIe-like links. Given that the electronic path is actually on FPGA, while some profiling uses GPU, the estimate should be rebuilt with a consistent hardware model, including I/O and thermal/settling times for reprogramming the mesh.

We thank the reviewer for this comment, which allows us to clarify the scope and rationale behind our latency analysis.

1) Consistency of the Hardware Model and Comparison Baseline: The latency of 49.5 ps/pixel is a calculated metric for the core computational pipeline of our OECP, encompassing the photonic matrix multiplication and the on-FPGA execution of the LiPICO-Net reconstruction algorithm. This is directly comparable to the computational latency of the AI models (ReconNet, DR2-Net) on a GPU, which is the standard and most relevant baseline for assessing the computational speed advantage of our photonic processor. The profiling of the baseline models on a GPU (Nvidia RTX 4090) is the appropriate and standard methodology, as these models are designed for and typically run on such hardware. Our comparison thus fairly evaluates the computational workload executed on two different platforms (our OECP vs. a GPU).

2) Focus on the Computational Core and Established Practice: The primary objective of our latency analysis is to isolate and highlight the transformative performance of the photonic computational core and its co-designed lightweight algorithm. As with the energy analysis (Response to Comment 4), it is standard practice in the photonic computing literature to report the intrinsic computational latency of the core operation, which provides a clear metric for the technology's potential. The data movement and serialization overheads mentioned by the reviewer are system-level implementation details that are highly dependent on the specific system architecture and are not part of the core algorithm's computational time on either platform (GPU or our OECP). Currently we're using commercially available products of

electronic components with low price and poor performance (due to research budget), including them would obscure the fundamental comparison of computational efficiency.

To prevent any misunderstanding, we will revise the manuscript to more explicitly state the scope of our latency calculation. We will clarify that the 49.5 ps/pixel metric represents the **core computational latency** for the compression and reconstruction algorithm, enabling a direct comparison with the computational latency of the baseline models on a GPU. We will also note that this demonstrates the intrinsic speed of the photonic-electronic co-design, and that further system-level integration (e.g., using ASICs instead of an FPGA for control) could be pursued in future work to minimize all ancillary overheads for specific applications.

6. The reconstruction is shown optically, while the compression stage remains electronic (as shown in Fig. 3), hence the claimed end-to-end photonic system is not demonstrated. A complete physical loop from optical encoding to reconstruction with real optics at both stages would materially strengthen the work.

We respectfully disagree that the compression stage remains electronic. Fig. 3(a) shows the whole compression-reconstruction process of our LiPICO-Net, both the compression and reconstruction part are **implemented on photonics** (except the fully connected layer). The **compression matrix is generated by the PIC chip and the matrix multiplication is done on the PIC chip**, so it should be proper to claim that the compression stage is photonic.

7. There is a discrepancy in a cited latency figure for Ref. [1] (authors give 3.91 ps/pixel, whereas the source implies ≈ 0.98 ps/pixel under stated assumptions). All third-party metrics must be recalculated faithfully to avoid misleading comparisons.

We appreciate the reminder regarding citation rigor, but we maintain the correctness of the listed data. The value of 0.98 ps/pixel **was calculated under the condition of $N=64$, as assumed in Ref. [1]**:

Assuming an image block size of 8×8 ($N = 64$), this approach can process a Terapixel/second of image data using a clock speed of ~ 16 GHz

While they **experimentally fabricated a photonics chip with $N=16$** :

we experimentally characterized a passive silicon photonic prototype with 16 inputs designed to process 4×4 pixel kernels

Since the actually fabricated chip has 16 inputs rather than 64, the time consumption per pixel should be four times that of the assumed scenario. We believe it should be fair to use the results from the actual device for all data comparisons.

Ref.:

1. Nature Communications volume 15, Article number: 4510 (2024)

8. A 32×32 optical core is small relative to megapixel imagery, please clarify tiling/streaming strategy, line-rate, buffer sizes, and throughput limits when scaling beyond toy patches.

We thank the reviewer for this important question regarding the scalability of our system for large-scale imagery. We provide the following clarifications:

1) A monolithic, programmable 32×32 photonic computing core (PCC) with integrated modulators and detectors, as demonstrated in our work, represents a scale that is currently at the forefront of integrated photonic computing systems (as pointed by other reviewers). For context, a very recent seminal work published in Nature (Volume 640, pages 361–367, 2025) demonstrated a 64×64 core, which is among the largest reported to date. Our 32×32 PCC is thus well within the state-of-the-art and provides a substantial and practical computational unit for real-world applications.

2) The methodology of partitioning a large image into smaller, non-overlapping patches (e.g., 32×32 pixels) for compressive sensing and subsequent reconstruction is a standard and well-established practice in the field. This approach is universally employed by classical CS reconstruction models, including ReconNet and DR2-Net, which our work builds upon. This tiling strategy is not a limitation but a fundamental and effective aspect of the algorithm that makes the processing of megapixel images tractable. Our system is designed to operate on this principle. When processing an image, we first divide it into non-overlapping 32×32 patches which are then processed individually by LiPICO-Net. Then the reconstructed patches are reassembled into their original positions to complete the processing of the entire image. Since the network always deals with 32×32 patches, the size of megapixel imagery does not affect the processing capability of the optical core. Quotes from [1]:

The scene is divided into non-overlapping blocks. Each block is reconstructed by feeding in the corresponding CS measurements to ‘ReconNet’.

3) Regarding the requested metrics. The system data flow is divided into optical and electrical paths:

i. Optical Path: Laser $\rightarrow$ Polarization Controller $\rightarrow$ Input Grating Coupler $\rightarrow$ 1×32 Splitter $\rightarrow$ 32-channel Modulators $\rightarrow$ Optical Network $\rightarrow$ 32-channel Photodetectors

ii. Electrical Path: FPGA $\rightarrow$ DAC $\rightarrow$ Operational Amplifier $\rightarrow$ Modulators $\rightarrow$ Optical Network $\rightarrow$ Photodetectors $\rightarrow$ TIA $\rightarrow$ ADC $\rightarrow$ FPGA

While the algorithmic data flow comprises:

i. DDR $\rightarrow$ Input Buffer $\rightarrow$ DAC Module $\rightarrow$ Electrical Path $\rightarrow$ Optical Path $\rightarrow$ Electrical Path $\rightarrow$ ADC Module $\rightarrow$ ADC Buffer $\rightarrow$ MAC Unit $\rightarrow$ Output Buffer $\rightarrow$ DDR

ii. A Kernel Buffer supplies the input to the MAC unit (primarily for quantization).

buffer sizes:

i. Input buffer: $224 \times 224 \times 8\text{bit}$

ii. ADC buffer: $112 \times 7 \times 7 \times 16\text{bit}$

iii. Output buffer: $56 \times 56 \times 256 \times 16\text{bit}$

iv. Kernel buffer: $2048 \times 16\text{bit}$

Throughput is limited by the FPGA's internal memory.

Ref:

1. K. Kulkarni, et al., "Reconnet: Non-iterative reconstruction of images from compressively sensed measurements," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016), 449-458.

9. Report MS-SSIM/LPIPS at high ratios e.g., $256\times$ where PSNR ≈ 17 dB is borderline useful. Provide visual error maps and task-level utility, such as detection/segmentation after reconstruction.

We thank the reviewer for raising these pertinent points. We have followed the comment to include MS-SSIM metrics at high ratios (e.g., $256\times$) and supplemented the evaluation with MS-SSIM and SSIM metrics for all images mentioned in the manuscript. We also provided the corresponding Viridis color maps of the actual photo.

Concerning the evaluation of task-level utility, we respectfully clarify that our present work concentrates on the fundamental task of image compression and reconstruction. While we recognize the growing interest in task-driven compression, a comprehensive evaluation across various downstream tasks (e.g., detection, segmentation) represents a significant separate research endeavor. As such, many comparable studies in the compression sensing domain also primarily focus on pixel-level and perceptual metrics. We hope the reviewer agrees that this established focus remains valid for demonstrating our contributions.

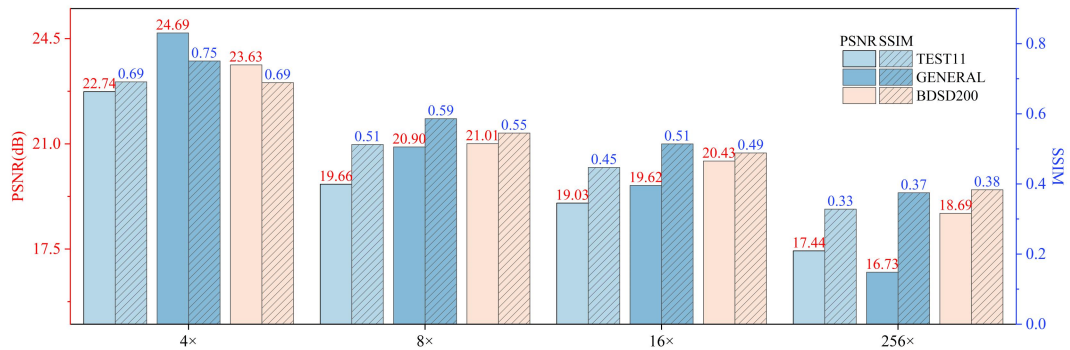


Fig. R1 Performance comparison of OECP performing end-to-end compression and reconstruction of multiple image datasets at various compression ratios.

Compression Ratio Data Category		4×	10×	25×	100×
LiPICO-Net(AE)	PSNR(dB)	25.48 ± 2.52	22.31 ± 2.14	20.28 ± 2.14	17.90 ± 1.96
	SSIM	0.7796 ± 0.0714	0.6339 ± 0.0875	0.5034 ± 0.0956	0.3717 ± 0.1206
	MS-SSIM	0.7705 ± 0.0745	0.6263 ± 0.0915	0.5052 ± 0.1025	0.3899 ± 0.1247
LiPICO-Net(RM)	PSNR(dB)	25.17 ± 2.34	22.28 ± 2.09	20.25 ± 2.13	17.79 ± 1.92
	SSIM	0.7637 ± 0.0705	0.6319 ± 0.0822	0.5049 ± 0.0929	0.3690 ± 0.1180
	MS-SSIM	0.7527 ± 0.0723	0.6236 ± 0.0853	0.5066 ± 0.0998	0.3868 ± 0.1222

Table R1 Mean and standard deviations of PSNR, SSIM and MS-SSIM of LiPICO-Net(AE) and LiPICO-Net(RM) on 11 images.

Compression Ratio Data Category		4×	8×	16×	256×
Actual Photo	PSNR(dB)	30.57	26.40	25.78	23.13
	SSIM	0.8463	0.7256	0.6699	0.5760
	MS-SSIM	0.8519	0.7480	0.7001	0.6241
Ultra-large Image	PSNR(dB)	34.00	30.35	28.71	26.56
	SSIM	0.9081	0.8191	0.7811	0.7133
	MS-SSIM	0.9124	0.8395	0.8089	0.7558

Table R2 Mean and standard deviations of PSNR, SSIM and MS-SSIM of actual photo and ultra-large image.

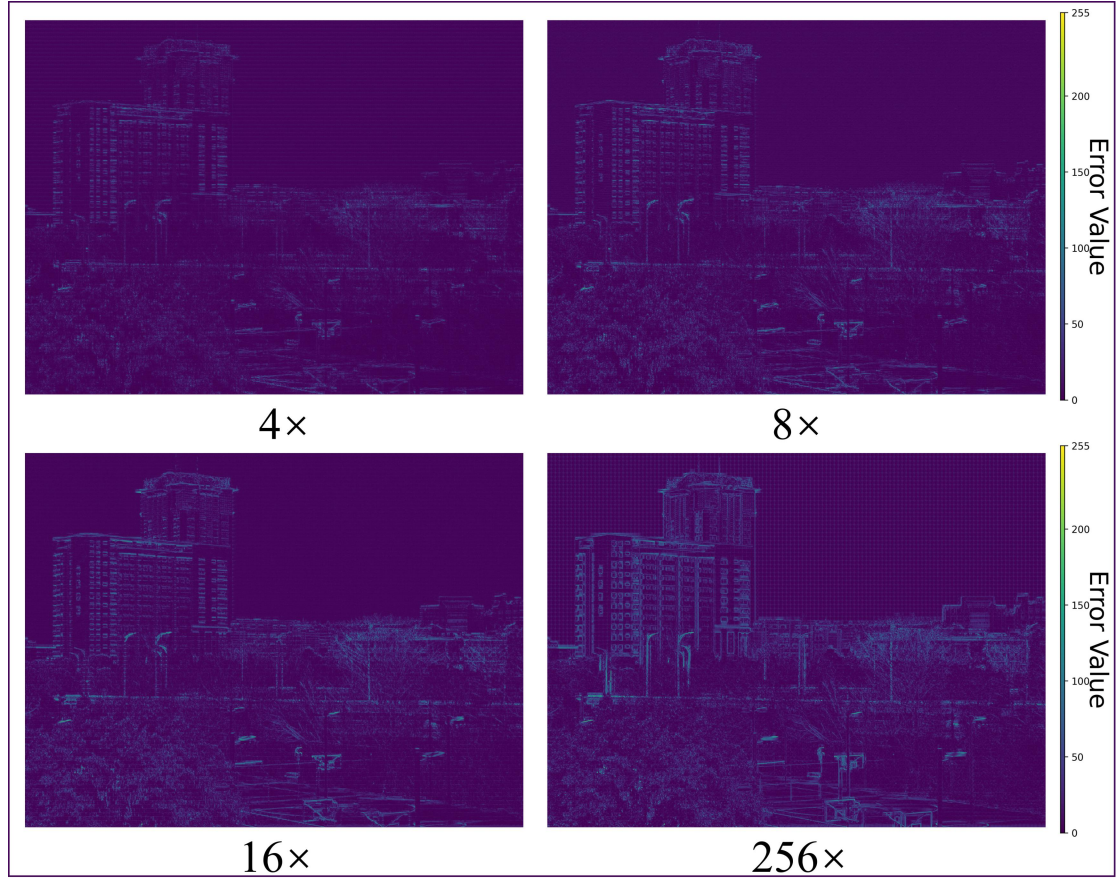


Fig. R2 Viridis color map of actual photo. The error visualization using the Viridis color map demonstrates the effectiveness of our reconstruction. The predominant dark purple hues across the image body indicate minimal reconstruction errors, while the sparse yellow regions are confined to fine details and edges, confirming good overall fidelity.

10. When reporting MAC/pixel reductions, ensure a hardware-consistent mapping, e.g., analog MAC equivalence, bit-precision, quantization, overhead of calibration and readout.

We thank the reviewer for raising this point regarding the characterization of computational operations. We respectfully offer the following perspective:

The primary objective of reporting Multiply-Accumulate operations per pixel (MAC/pixel) in this context is to provide a **platform-agnostic measure of the algorithmic complexity and intrinsic efficiency of the neural network model itself**. This established practice allows for a direct comparison of the computational load imposed by different network architectures (e.g., LiPICO-Net vs. ReconNet vs. DR2-Net), independent of their eventual hardware implementation. It is a standard and highly relevant metric in the machine learning and computational imaging communities for assessing model lightweights before deployment on any specific processing unit, be it a GPU, FPGA, or a photonic processor. Focusing on this algorithmic-level complexity is crucial for a fair comparison, as it isolates the gains achieved by our co-designed, photonic-hardware-aware model architecture. The reported >95%

reduction in MAC/pixel fundamentally demonstrates that LiPICO-Net is inherently a much more efficient network, which is a prerequisite for achieving low latency and low power consumption on any computing platform, including our photonic system.

But for further address the reviewer's concern, we'd like to provide additional discussions on the metrics required by the reviewer: The transmission property of the photonic chip is mathematically equivalent to performing a matrix multiplication between its transmission matrix and the input optical signal. Since most Multiply-Accumulate (MAC) operations can be formulated as matrix multiplications, the photonic chip is, in principle, equivalent to the simulated MAC operation. In practice, however, inherent analog imperfections in the chip and signals introduce errors. These become particularly pronounced for smaller elements in computations with a wide dynamic range, where they suffer from a substantially larger relative error.

We use a 14-bit DAC and a 12-bit ADC. However, due to errors introduced primarily by the analog electronic sections, the acquired data from the ADC has an effective resolution of approximately 6 bits.

The data acquired by the ADC is quantized to 16-bit. The intermediate temporary results of the multiply-accumulate (MAC) operations have a maximum bit width of 32 bits, while the final output of each layer is quantized to 8-bit.

For calibration, we employ four techniques:

- i. Applying an AC input signal to mitigate the impact of DC noise and drift.
- ii. Taking the average of multiple measurements while discarding the first result
- iii. Amplifying the input signal and correspondingly attenuating the output
- iv. Transfer Learning (retraining the other parts of the network under the condition of introducing errors).

The readout module and output buffer are physically integrated, which can be considered as having no readout overhead, or in other words, makes it difficult to isolate and measure the readout overhead separately.

11. Numerous typos/grammar issues such as “computtaional”, “haven’t to retrain total parameters”, unit typos “10.58 pJ/pixel” vs “nJ/pixel”, should be corrected, figures need axes/units and consistent legends.

We're sorry for those careless mistakes and we have gone through the manuscript twice to make sure those are properly corrected. However, the use of pJ and nJ does not represent unit typos, but rather the selection of appropriate units for energy consumption metrics that span orders of magnitude. The sources of these data can be found in the supplementary material.

In summary, the manuscript demonstrates engineering effort but, as written, lacks methodological novelty, fair evaluation, and credible system-level accounting commensurate with Nature Communications. A substantially revised version with like-for-like baselines, full system demonstration, rigorous power/latency tables, and rate-distortion evaluation might be suitable for a more specialized venue.

Once again, we appreciate the reviewer's thorough evaluation of our manuscript and many constructive feedback. We believe the reviewer has missed some contents in the manuscript and misunderstood the core of our work. We hope our responses can further clarify the novelty and value of our work and re-gained the recommendation from the reviewer.

Reviewer #2 (Remarks to the Author):

The authors demonstrate a programmable 32 by 32 linear operator for optical signal transformation. They demonstrate that the latency per input signal is less than 50 ps and an encoding power per pixel of 10 pJ. High speed tensor operations implemented by this chip could be used for many applications, here the authors choose to focus on image data compression and decompression. This is a powerful application for this device, it is particularly interesting that the authors here demonstrate the ability to load reconstruction weights on to the device. End to end, this enables a significant energy per pixel reduction relative to the passive encoding strategy implemented in their reference 28. As noted in the manuscript, flexible and adaptive compression ratios are useful in these applications. While one could consider numerous variations on how this photonic tensor operator might be applied, the approach demonstrated here is sound, the paper is well written and the results are interesting. This work is of particular interest because it is an end-to-end demonstration of very practical application of photonic neural processing.

We are truly delighted and honoured to receive such positive and insightful comments from the reviewer. We sincerely appreciate the reviewer's recognition of our work, especially the remarks that our system represents "a powerful application" and "a programmable 32 by 32 linear operator for optical signal transformation" with "latency per input signal of less than 50 ps and an encoding power per pixel of 10 pJ." It is particularly encouraging that the reviewer highlighted the significance of loading reconstruction weights on the device, noting that this enables "a significant energy per pixel reduction relative to the passive encoding strategy." We are also grateful for the reviewer's observation that our approach is "sound," the paper is "well written," and the results are "interesting." Most of all, we thank the reviewer for emphasizing that "this work is of particular interest because it is an end-to-end demonstration of very practical application of photonic neural processing." These thoughtful comments affirm the value of our contributions and motivate us to continue advancing this line of research.

Reviewer #3 (Remarks to the Author):

This work presents an image compression and reconstruction scheme based on a programmable photonic processor. The fundamental part of the work is very similar to the previously published study (X. Wang et al., “Integrated photonic encoder for low-power and high-speed image processing,” Nature Communications 15, 4510 (2024)). As a novelty, the authors provide two new contributions:

- First, the authors propose and experimentally demonstrate the use of a 32×32 programmable mesh based on Mach–Zehnder interferometers with an FFT architecture to implement random matrix multiplication, which presents advantages in terms of reconfigurability when compared to random scattering devices such as those previously demonstrated.
- Second, the authors use an alternative neural network algorithm for the reconstruction of the encoded images, which provides comparable results to more complex state-of-the-art solutions.

We appreciate Reviewer 3’s overall positive evaluation of our work and have carefully revised our manuscript based on reviewer’s insightful comments.

While the work is of interest to the community and is clearly written, I cannot recommend the publication of the article in its current form, as the manuscript omits some important points and includes some misleading statements, especially those related to the photonic processor and its capabilities:

- Throughout the article, the authors claim several times that this is the first demonstration of a compression and reconstruction approach using an optoelectronic computing processor. However, as they themselves mention, a similar work was previously published, which is also optoelectronic. The authors should revise this statement.

We respectfully disagree that our statement should be revised. Indeed, Ref. 1 demonstrated an photonic image compression scheme, where the image compression is purely performed optically using a passive photonic device, without any electronics. While the reconstruction part is completely implemented on electronics (PC I assume), making the whole system not able to be called optoelectronic. While our compression and reconstruction part both utilize our photonics computing chip, we think it’s proper to call it the first demonstration of a compression and reconstruction approach using an optoelectronic computing processor.

- Regarding the novelty of the work, the authors focus their comparison on the differences with X. Wang et al., “Integrated photonic encoder for low-power and high-speed image processing,” Nature Communications 15, 4510 (2024). In the photonics domain, the fundamental difference lies in the programmability of the photonic processor. The authors propose the use of an FFT-based architecture, which is less complex than other solutions such as the rectangular (Clements) or triangular (Reck) configurations, at the expense of reduced matrix accessibility. I wonder whether this limitation in the number of matrices that can be implemented may represent a fundamental restriction for achieving better results. A comparison with other photonic architectures should be included in this regard. Moreover, the

authors should provide more information about the architecture used, as well as its benefits and limitations.

We thank the reviewer for this insightful comment regarding the choice of our photonic architecture and its potential impact on performance. The reviewer correctly identifies a key design trade-off. It is true that the FFT-based mesh, while more compact and lower-loss, does not possess the universal representation capabilities of the Clements or Reck architectures. This means it cannot implement every possible unitary matrix, which theoretically could limit the expressiveness of the compressive matrices it generates. However, we would like to clarify that this was a deliberate and advantageous engineering choice for the following reasons:

1) Our FFT-based mesh requires only 80 MZIs, a significant reduction compared to the ~ 500 MZIs needed for a full 32×32 Clements mesh. This directly translates to fewer thermo-optic phase shifters (192 in our design), drastically lowering the static power consumption for configuring the chip and simplifying control electronics. Fewer MZI also indicates lower optical losses, which is crucial for maintaining a high signal-to-noise ratio (SNR) and thus the computational fidelity of the system. Besides, as supported by prior research (e.g., M. Y.-S. Fang et al., *Optics Express* 27, 14009, 2019), the FFT topology demonstrates greater robustness to component imperfections and phase errors compared to the Clements structure. This higher intrinsic tolerance was a decisive factor for achieving high yield and performance in a practical, fabricated system.

2) The perceived limitation in matrix accessibility is effectively mitigated by our end-to-end co-design framework. During the training of our system, the compression matrix (programmed into the PCC) and the reconstruction network (LiPICO-Net) are co-optimized. The training algorithm naturally learns to find highly effective compressive matrices that are within the expressive range of our FFT-based PCC. **It does not need an arbitrary matrix; it needs a good matrix that works in concert with LiPICO-Net.**

3) The practical consequence of this architectural choice is remarkably small. The performance gap between the ideal simulation (LiPICO-Net (AE)) and the physically implemented system (LiPICO-Net (RM)) in Fig. 3(b) is only on the order of a few tenths of a dB in PSNR. This minor degradation confirms that the FFT mesh provides a sufficiently rich set of matrices to achieve high-fidelity compression and reconstruction, while granting us the substantial hardware advantages listed above

To further address the reviewer's concern, we show the accuracy of different architectures in expressing Unitary Matrix, Pseudo-Real-Value Matrix and SVD Matrix.

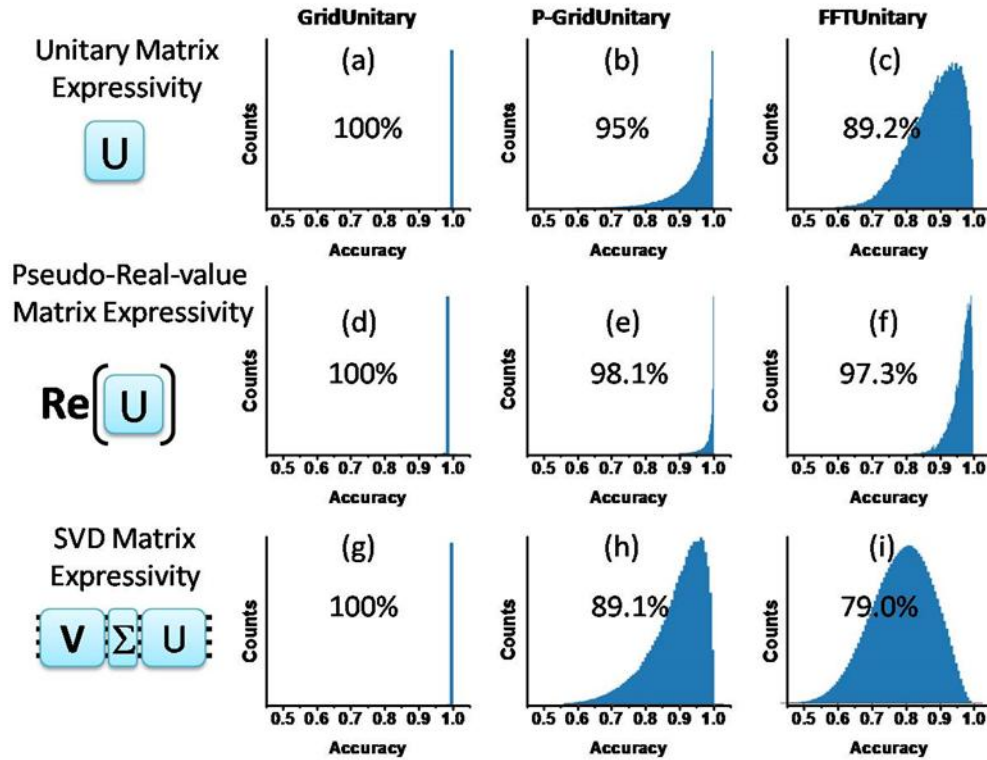


Fig. R1 The accuracy of different architectures in expressing Unitary Matrix, Pseudo-Real-Value Matrix and SVD Matrix. from reference [1]

Ref:

1. Tian, Ye, et al. "Scalable and compact photonic neural chip with low learning-capability-loss." *Nanophotonics* 11.2 (2022): 329-344.

- Regarding the photonic processor:

o The authors mention the inclusion of 32 high-speed modulators and photodetectors that allow a working speed of more than 30 GHz. They also provide the experimental response of these devices. During the experiments, which signal frequency was used? In case lower-frequency signals were employed, the authors should adapt their metrics (energy/pixel, latency, etc.) and present both the experimental and the achievable figures.

We first give the signal frequency required by the reviewer: The ADC sampling rate used in our system is 2 GSPS, and the DAC operates at 6 GSPS, meaning the complete computing board can support a computational frequency of 1 GHz. Then we clarify why the performance of electronics shall not be included in the overall performance metrics:

1) The primary contribution of our work is the demonstration of a monolithically integrated photonic core (PCC) that performs the computationally intensive matrix multiplication at the speed of light with ultra-low latency and power. The standard and most critical metric for evaluating such a photonic computing core is its intrinsic performance – the energy and latency for the fundamental optical operation (matrix-vector multiplication), as this directly

showcases the photonic advantage. This approach of reporting the core photonic processor's performance is well-established in the field [e.g., Refs. 17, 21, 22, 27, 28 in our manuscript], as it allows for a direct comparison of the photonic technology itself, independent of the peripheral electronics which can be optimized separately and follow their own Moore's Law scaling.

2) We emphasize that the 30-40 GHz bandwidths measured for the modulators and detectors (Figs. 4d,e) represent the component-level speed capability, demonstrating the platform's potential for extreme throughput. The actual operating speed and thus the power of the ADCs/DACs in our current proof-of-concept demonstration are dictated by the FPGA and the image processing pipeline, not the ultimate photonic bandwidth. Future iterations of this work, aimed at maximizing throughput, would indeed require co-design with advanced ASICs for the electronics.

o Little information is provided regarding the photonic building blocks, which reduces the reproducibility of the work. System insertion losses (MZIs, optical packaging, etc.) should be reported.

We agree that the building block should be introduced with more details and apologize for the missing information. The insertion loss of MZI is approximately 0.1 dB and the insertion loss of grating coupler is approximately 4 dB. The waveguide propagation loss and waveguide crossing loss are 1.80 dB/cm and 0.029 dB, respectively. The grating coupler has a 3 dB bandwidth of 60 nm and a 1 dB bandwidth of 35 nm. The modulator efficiency ($V\pi \cdot L\pi$) is 2 V·cm at -2 V bias with a 1.5 mm doping region length.

o The authors state that the photonic chip is co-packaged with the necessary control electronics, but I do not see the electronic elements in Fig. a–b. This statement, which is presented as a novelty, should be revised if the control electronics are independent from the optical processor.

We thank the reviewer for noting this oversight. We have revised the figure to include a photograph of the computing board with the integrated control electronics. And the direct image is shown in Fig.2(a)(b).

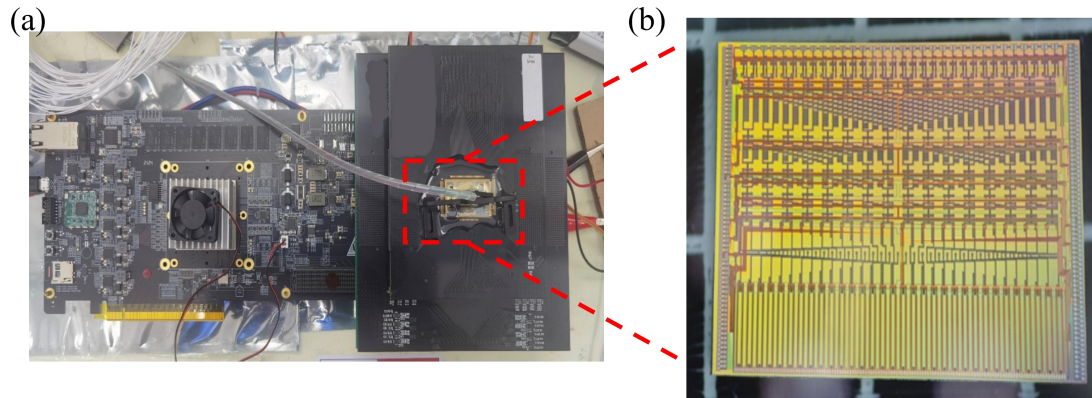


Fig. R2 Photos of the opto-electronic computing processor (OECF) (a) and the 32×32 photonic computing chip (PCC) (b), respectively.

o A 32×32 optical processor is at the edge of state-of-the-art photonic systems; however, it would be beneficial to include a scalability study of the photonic processor.

We agree that scalability is an important issue. Without significant improvement upon current fabrication technology of photonic chip, the FFT architecture currently achieves a maximum scale of 64×64 , limited by physical constraints including optical attenuation and thermal noise. Further scaling the matrix dimensions will necessitate more complex and costly approaches, such as multi-chiplet architectures or custom ASIC fabrication. This context is also included in the manuscript.

- Regarding LiPICO-Net, as I understand, it was adapted to work with the 32×32 size of the processor. How does it scale with the size of the input image?

This comment is similar with comment #8 of reviewer 1. When processing an image, we first divide it into non-overlapping 32×32 patches which are then processed individually by LiPICO-Net. Then the reconstructed patches are reassembled into their original positions to complete the processing of the entire image. Since the network always deals with 32×32 patches, no scaling based on the size of input image is required.

Please be noted that processing large-scale imagery by dividing it into blocks is a standard practice in the field of image processing[1]:

The scene is divided into non-overlapping blocks. Each block is reconstructed by feeding in the corresponding CS measurements to ‘ReconNet’.

Therefore, the size of the input image does not impose constraints on our system's processing capability.

Ref:

1. K. Kulkarni, et al., "Reconnet: Non-iterative reconstruction of images from compressively sensed measurements," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016), 449-458.

- Finally, concerning the estimation of performance metrics in the Supplementary Material, the authors mention that the energy contribution of the electronics such as ADC/DAC/TIA is negligible. However, the electrical part is usually the most power-consuming in photonic processors

(<https://pubs.aip.org/aip/app/article/7/2/020902/2835121/Scaling-up-silicon-photonic-based-accelerators>, <https://www.nature.com/articles/s41586-025-08854-x>). The authors should include the impact of the electronic part and adapt their comparisons accordingly.

We thank the reviewer for raising this critical point regarding energy accounting. We acknowledge that a comprehensive, component-level power breakdown is ideal for a full system-level comparison. However, we would like to clarify that the standard practices in the emerging field of photonic computing do not include the energy consumption of electronics and primarily focus on the consumption of photonic chips. Here are some concrete examples of performance characterization for PCC:

Shen et al. [1] calculate only the energy consumption of resistances on phase modulators:

Once all parameters have been trained and programmed on the nanophotonic processor, forward-propagation computing is performed optically on a passive system. In our implementation, maintaining the phase modulator settings requires some (small) power of ~ 10 mW per modulator, on average.

Xu et al. [2] only take the energy consumption of frequency comb they used into consideration:

The soliton crystal microcomb provided more than 90 channels over around 36 nm of the telecommunications C-band (1,540–1,570 nm), offering adiabatically generated low-noise frequency comb lines with a small footprint of <1 mm² and potentially low power consumption (<100 mW using the technique in ref. 31).

Feldmann et al. [3] consider only optical energy contributions:

A key feature of our approach is that, because the convolutional operation is a passive transmission measurement, the calculations can in theory be performed at the speed of light at very low power (17 fJ per MAC, considering only optical contributions)

Similarly, References [4-6] also accounted for only the optical energy consumption.

Besides, the pioneering work in [7], which is the first study to use photonic chip for image compression, doesn't incorporate the contribution from necessary electronics/computers when claiming "exceeding Terapixel/second using less than 100 fJ/pixel" and "this technique has

the potential to encode Terapixel/second data streams utilizing <100 femtojoules per pixel—representing a > 1000x reduction in power consumption compared with the state-of-the-art electronic approaches”, either. Since we establish our work on the foundation of this pioneer work, we believe it’s better to use the same standard.

As our work is performed on the basis of previous photonic computing chips as well as the pioneer work reported in [7], we believe it is proper to stick to the same criteria in order to provide a clear comparison.

But we have included additional discussions in the Supplementary Information for address the reviewer’s concern.

Ref:

1. Nature photonics 11, 441-446 (2017).
2. Nature 589, 44–51 (2021).
3. Nature 589, 52–58 (2021).
4. Nat Commun 14, 70 (2023).
5. Nature 569, 208-214 (2019).
6. Nature communications 13, 2131 (2022).
7. Nature Communications volume 15, Article number: 4510 (2024)

- Minor: There is a typo in the abstract where the authors write “computational.”

We’re sorry for those careless mistakes and we have gone through the manuscript twice to make sure those are properly corrected.

Reviewer #4 (Remarks to the Author):

The manuscript presents a significant advancement in photonic computing applied to image processing. The Authors demonstrate, for the first time, an end-to-end image compression and reconstruction system based on a fully integrated opto-electronic computing processor (OECF), achieving latency and energy efficiency improvements of several orders of magnitude compared with conventional GPU-based methods. The work is ambitious, timely, and highly relevant to the fields of silicon photonics, optical computing, and computational imaging. The key novelty lies in the co-integration of a 32×32 programmable photonic matrix core with high-speed modulators, detectors, and control electronics, enabling both compression and reconstruction in a unified photonic–electronic pipeline. The development of the lightweight LiPICO-Net, specifically optimized for photonic hardware constraints, further

enhances the originality. Compared to prior work such as Wang et al. [28], this study achieves adjustable compression ratios and full integration, which constitute major improvements.

We appreciate Reviewer 4's overall positive evaluation of our work and have carefully revised our manuscript based on reviewer's professional comments.

Here, my major comments:

1. The methodology is carefully described in both the main text and supplementary materials. The description of LiPICO-Net (Supplementary S1) is comprehensive, including the convolutional architecture and mathematical formulation, while the details of the opto-electronic hardware (Supplementary S2) and performance estimations (S3) are clearly presented. However, several methodological aspects would benefit from further elaboration to reinforce the technical soundness of the work.

a. The reported latency and energy consumption values rely on idealized assumptions regarding the modulation bandwidth, optical power levels, and overall system clock rate. It is not entirely clear how these parameters were experimentally verified or constrained by measurement. For instance, while the modulator and detector bandwidths are reported as exceeding 30 and 40 GHz respectively, it would be valuable to discuss whether the 30 GHz operational frequency (used to estimate the MAC time of 0.032 ps) represents a measured limit or a theoretical upper bound. A brief uncertainty analysis or sensitivity assessment—quantifying how variations in optical power or bandwidth impact latency and energy per pixel—would significantly strengthen the reliability of the reported performance metrics.

We thank the reviewer for the suggestion. Our optical network (including the PCC, modulators, and photodetectors) has experimentally achieved an operational frequency of 30 GHz, which is limited by the bandwidth of modulators and photodetectors. But, the full system currently cannot operate at 30 GHz, owing to limitations of the performance of commercially available ADCs, DACs and FPGA. Due to research budget, we were unable to utilize state-of-the-art commercial ADCs, DACs, and other electronic components. The ADC sampling rate used in our system is 2 GSPS, and the DAC operates at 6 GSPS, meaning the complete computing board can support a computational frequency of 1 GHz. However, we believe this does not weaken the value of our work. With the ongoing development of ADC and DAC devices, or through potential future alternatives for digital-to-analog conversion, the high-speed performance of our optical network can be fully realized.

Per the reviewer's request, we perform an analysis of how the latency per pixel and energy per pixel vary with power and bandwidth. As calculated in Section 2.1 of Supplementary Information, the latency per pixel is only related to the bandwidth, specifically:

$$Latency = \tau_0 + \frac{M_r}{M_0 f_0} + t_{FC}$$

where f_0 is the bandwidth, M_0 is the number of MAC per calculation, M_r is the MAC per pixel in reconstruction, and t_{FC} is the latency of FCL.

As calculated in Section 2.2 of Supplementary Information, the energy per pixel is related to power and bandwidth, specifically:

$$Energy = \frac{(M_c + M_r) \times (E_0 + E_e)}{M_0} + E_{FC}$$

where M_c is the MAC per pixel in compression, E_{FC} is the energy of FCL.

b. The validation of the reconstruction quality relies exclusively on the Peak Signal-to-Noise Ratio (PSNR). While PSNR remains a standard metric, it often fails to capture perceptual fidelity or structural accuracy in complex images. Including additional metrics such as Structural Similarity Index (SSIM) or Learned Perceptual Image Patch Similarity (LPIPS) would provide a more robust and perceptually relevant evaluation of the reconstruction quality, particularly at high compression ratios where visual artifacts may dominate.

We thank the reviewer for raising these pertinent points. We understand that PSNR alone may not fully capture reconstruction performance. Therefore, we have supplemented the evaluation with MS-SSIM and SSIM metrics for all images mentioned in the manuscript, which is also recommended by reviewer 1.

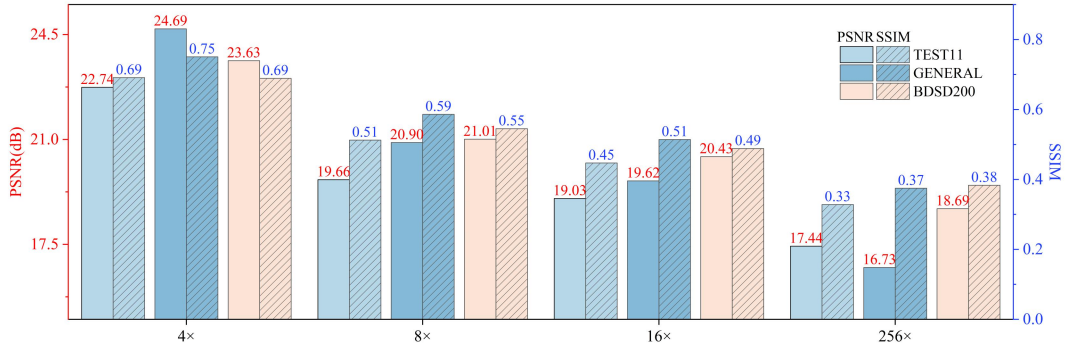


Fig. R1 Performance comparison of OECP performing end-to-end compression and reconstruction of multiple image datasets at various compression ratios.

Compression Ratio Data Category		4×	10×	25×	100×
LiPICO-Net(AE)	PSNR(dB)	25.48 ± 2.52	22.31 ± 2.14	20.28 ± 2.14	17.90 ± 1.96
	SSIM	0.7796 ± 0.0714	0.6339 ± 0.0875	0.5034 ± 0.0956	0.3717 ± 0.1206
	MS-SSIM	0.7705 ± 0.0745	0.6263 ± 0.0915	0.5052 ± 0.1025	0.3899 ± 0.1247
LiPICO-Net(RM)	PSNR(dB)	25.17 ± 2.34	22.28 ± 2.09	20.25 ± 2.13	17.79 ± 1.92
	SSIM	0.7637 ± 0.0705	0.6319 ± 0.0822	0.5049 ± 0.0929	0.3690 ± 0.1180
	MS-SSIM	0.7527 ± 0.0723	0.6236 ± 0.0853	0.5066 ± 0.0998	0.3868 ± 0.1222

Table R1 Mean and standard deviations of PSNR, SSIM and MS-SSIM of LiPICO-Net(AE) and LiPICO-Net(RM) on 11 images.

Compression Ratio Data Category		4×	8×	16×	256×
Actual Photo	PSNR(dB)	30.57	26.40	25.78	23.13
	SSIM	0.8463	0.7256	0.6699	0.5760
	MS-SSIM	0.8519	0.7480	0.7001	0.6241
Ultra-large Image	PSNR(dB)	34.00	30.35	28.71	26.56
	SSIM	0.9081	0.8191	0.7811	0.7133
	MS-SSIM	0.9124	0.8395	0.8089	0.7558

Table R2 Mean and standard deviations of PSNR, SSIM and MS-SSIM of actual photo and ultra-large image.

c. The reported total energy consumption per pixel appears to primarily account for the photonic matrix operations. However, the OECP includes multiple electronic components—FPGA, ADCs, DACs, and TIAs—that can contribute significantly to total energy consumption, especially in high-speed operation. It should be clarified whether these electronic contributions were measured, estimated, or neglected in the presented calculations. If omitted, the authors should specify this explicitly and discuss how their inclusion would affect the overall system-level efficiency.

We thank the reviewer for raising this critical point regarding energy accounting. In our manuscript, we only include the energy consumption of the photonic computing chip as well as the full-connection-layer executed by the FPGA. We would like to clarify that the standard practices in the emerging field of photonic computing do not include the energy consumption of electronics and primarily focus on the consumption of photonic chips. Here are some concrete examples of performance characterization for PCC:

Shen et al. [1] calculate only the energy consumption of resistances on phase modulators:

Once all parameters have been trained and programmed on the nanophotonic processor, forward-propagation computing is performed optically on a passive system. In our implementation, maintaining the phase modulator settings requires some (small) power of ~10 mW per modulator, on average.

Xu et al. [2] only take the energy consumption of frequency comb they used into consideration:

The soliton crystal microcomb provided more than 90 channels over around 36 nm of the telecommunications C-band (1,540–1,570 nm), offering adiabatically generated low-noise frequency comb lines with a small footprint of <1 mm² and potentially low power consumption (<100 mW using the technique in ref. 31).

Feldmann et al. [3] consider only optical energy contributions:

A key feature of our approach is that, because the convolutional operation is a passive transmission measurement, the calculations can in theory be performed at the speed of light at very low power (17 fJ per MAC, considering only optical contributions)

Similarly, References [4-6] also accounted for only the optical energy consumption.

Besides, the pioneering work in [7], which is the first study to use photonic chip for image compression, doesn't incorporate the contribution from necessary electronics/computers when claiming "exceeding Terapixel/second using less than 100 fJ/pixel" and "this technique has the potential to encode Terapixel/second data streams utilizing <100 femtojoules per pixel—representing a > 1000x reduction in power consumption compared with the state-of-the-art electronic approaches", either. Since we establish our work on the foundation of this pioneer work, we believe it's better to use the same standard.

As our work is performed on the basis of previous photonic computing chips as well as the pioneer work reported in [7], we believe it is proper to stick to the same criteria in order to provide a clear comparison.

But we have included additional discussions in the Supplementary Information for address the reviewer's concern.

Ref:

1. Nature photonics 11, 441-446 (2017).
2. Nature 589, 44–51 (2021).
3. Nature 589, 52–58 (2021).
4. Nat Commun 14, 70 (2023).
5. Nature 569, 208-214 (2019).

6. Nature communications 13, 2131 (2022).
7. Nature Communications volume 15, Article number: 4510 (2024)

2. The manuscript frequently refers to “computing” and “computation” in the context of photonic matrix multiplications. While this is technically correct, it would be important to stress that the scope of photonic computing extends well beyond linear algebraic operations. In the broader context of optical information processing, “computing” encompasses a wide range of primitives, including discrete and continuous Fourier transforms (FFT), convolutions, correlations, and nonlinear mappings, which are increasingly realized through integrated photonic architectures (see, for example, doi: 10.1109/JLT.2024.3453670). By acknowledging this broader framework, the authors could more clearly position their work within the expanding field of hybrid analog–digital photonic computing, highlighting how the proposed OECP architecture might be adapted to more general-purpose or signal-processing-oriented tasks in the future.

We appreciate the reviewer’s professional suggestion. The context in the manuscript is modified considering reviewer’s suggestion and the relevant reference has also been included and properly acknowledged.

3. The performance results are impressive and generally well supported by quantitative data. The comparison in Table 1 effectively summarizes the advantages over GPU-based models, both in latency and energy efficiency. Nevertheless, a deeper and more critical discussion of scalability and practical implementation limits would improve the manuscript.

a. Scalability represents a key challenge for photonic matrix processors. As the number of optical channels increases beyond 32×32 , phase stability, thermal crosstalk, and calibration complexity can grow nonlinearly. The authors should discuss to what extent the FFT-based mesh architecture remains practical for larger matrices and whether emerging alternatives—such as modular photonic tensor cores or folded mesh layouts—might offer improved scalability.

We agree that scalability is an important issue. Without significant improvement upon current fabrication technology of photonic chip, the FFT architecture currently achieves a maximum scale of 64×64 , limited by physical constraints including optical attenuation and thermal noise. Further scaling the matrix dimensions will necessitate more complex and costly approaches, such as multi-chiplet architectures or custom ASIC fabrication. This context is also included in the manuscript.

b. The manuscript should clarify how system performance scales with compression ratio, particularly near the extreme regime of $256\times$. At such high ratios, optical noise, quantization effects, and the conditioning of the compression matrix may significantly affect

reconstruction quality. Quantitative results on PSNR degradation or convergence behavior during training across different ratios would provide valuable insight into system robustness.

We thank the reviewer for helpful comment and professional insight. To analyse quantitative results on PSNR degradation or convergence behaviour, we utilize LiPICO-Net(AE) which can easily achieve accurate compression ratios and have accordingly expanded our tests with more compression ratios to generate quantitative results of degradation. Considering the reviewer's comments below, we provide the PSNR, SSIM and MS-SSIM results for comprehensive evaluation.

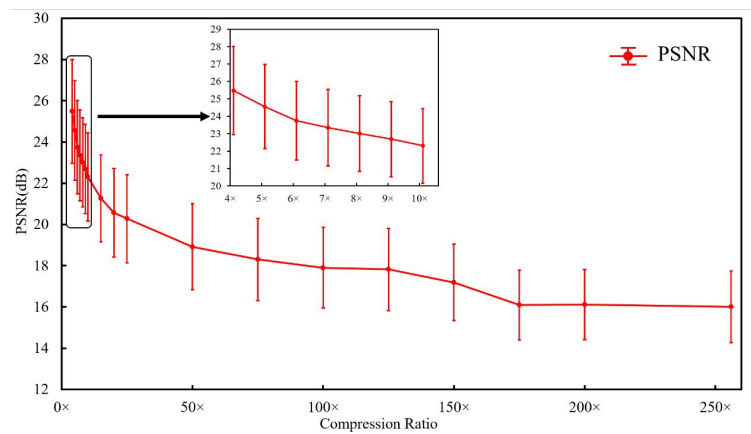


Fig. R1 Average PSNR versus Compression Ratio of 11 images.

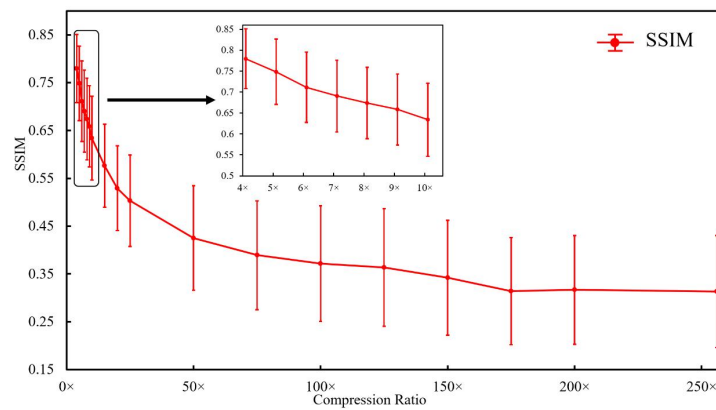


Fig. R2 Average SSIM versus Compression Ratio of 11 images.

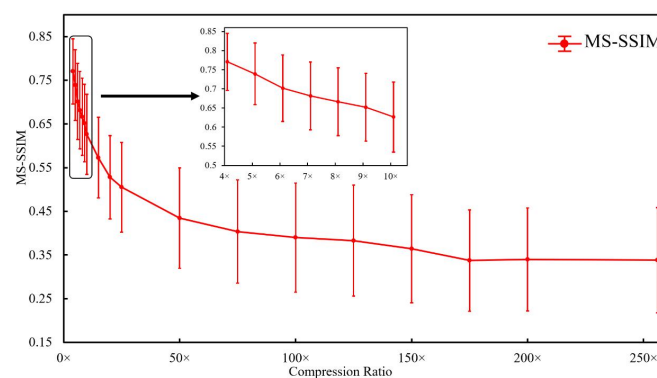


Fig. R3 Average MS-SSIM versus Compression Ratio of 11 images.

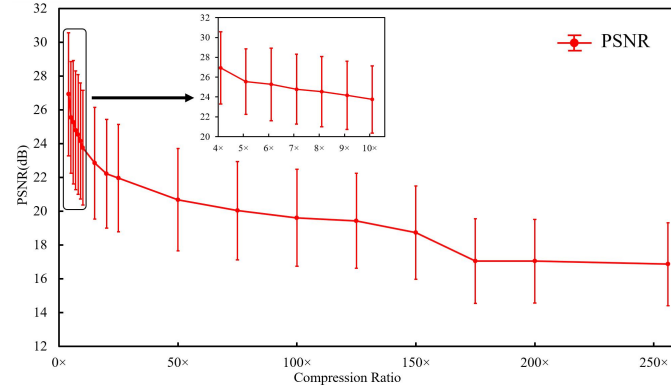


Fig. R4 Average PSNR versus Compression Ratio of BSDS 200.

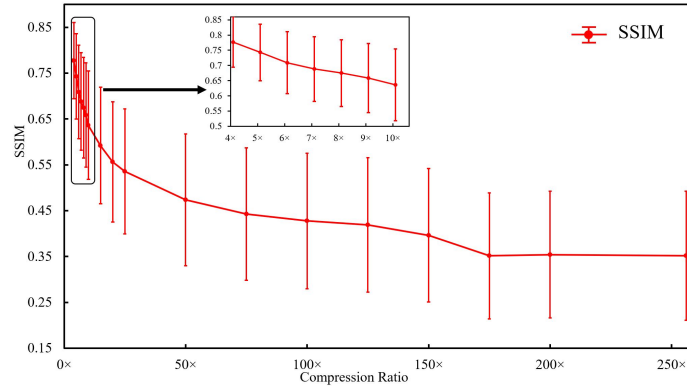


Fig. R5 Average SSIM versus Compression Ratio of BSDS 200.

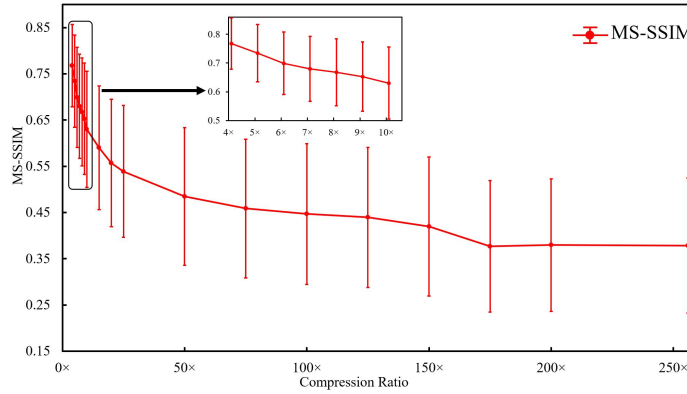


Fig. R6 Average MS-SSIM versus Compression Ratio of BSDS 200.

c. The degradation observed between simulated (AE) and experimental (RM) results in Fig. 3(b) deserves more attention. It would be helpful to identify the dominant physical sources of this discrepancy—such as phase noise, coupling losses, MZI imbalances, or limited precision in the programmable phase shifters—and to discuss potential mitigation strategies, including calibration protocols or feedback control.

Indeed, it is anticipated a discrepancy between a simulation results and experimental results, and we're glad that the discrepancy here is relatively small (limited to less than 0.5dB). The results of simulated (AE) are based on arbitrary matrices generated by computer, while the results of experimental (RM) are based on physical matrices generated by PCC. The theoretical reason for degradation is that the FFT architecture of PPC is incomplete and cannot express arbitrary matrices, but as in the response to previous reviewer, using FFT architecture can bring some noticeable advantages in terms of energy consumption, computational precision as well as robustness to fabrication imperfections.

Physical reasons for degradation lie in two parts:

1. Chip-Inherent Errors:

- i. The 1×16 beam splitter fails to achieve perfect power uniformity across outputs, requiring compensation through individual modulator adjustments to balance optical power in each path.
- ii. Inconsistent insertion loss and non-50:50 splitting ratios among different MZI beam splitters prevent perfect matrix representation. We employ blind calibration to approximate target matrices by iteratively tuning all MZI phase shifter voltages to minimize the discrepancy between measured and ideal outputs.
- iii. Non-uniform photodetector responsivity necessitates post-ADC scaling factors to normalize signal magnitudes across different detection channels.
- iv. Thermal drift and mechanical vibration affect optical properties, requiring active temperature stabilization and minimal environmental disturbance during operation.

2. Control Board Errors:

- i. Signal crosstalk between modulators and detectors through silicon interposers constitutes a yield-related issue, mandating the disposal of affected interposer units.
- ii. Analog signal chain limitations in AD/DA conversion introduce SNR constraints. We employ operational amplifiers for signal boosting, AC coupling with bandpass filtering to suppress DC offset and white noise, and multi-sample averaging for noise reduction.

4. The figures are generally clear, well organized, and of high visual quality. Figure 4(g) provides a convincing illustration of the reconstruction performance at different compression ratios. Nonetheless, some improvements to data presentation would increase the transparency of the results.

- a. Each figure caption should clearly distinguish between simulated data, idealized results, and experimentally measured quantities. For instance, specifying “simulation (AE)” or “experimental (RM)” in the captions themselves would help readers interpret the data correctly without needing to refer back to the text.

Thanks for the suggestion. We have revised the figure captions to make them more reader-friendly.

b. Axis labels and units in all quantitative plots, particularly those showing PSNR or latency versus compression ratio, should be explicitly stated to ensure reproducibility and interpretability.

We agree with the reviewer and have updated all quantitative plots with clear axis labels and units to ensure reproducibility and interpretability.

5. The supplementary information (Sections S1–S3) is well structured and provides essential implementation details. The mathematical derivation of LiPICO-Net and the step-by-step energy and latency estimations are valuable. However, cross-referencing between the main manuscript and supplementary materials could be made more systematic. For example, when equations or design parameters are derived in the supplementary sections, the main text should explicitly refer to the corresponding section or equation number. This would enhance readability and help readers follow the logical connection between modeling assumptions and experimental implementation.

Thanks for the suggestion. We have gone through the manuscript twice and included additional cross-reference between main context and Supplementary Information.

6. The manuscript is overall well written and logically structured. The exposition is clear, and the technical depth is appropriate for a high-impact journal. However, several stylistic and minor linguistic improvements are recommended. A few typographical errors persist (e.g., “computtaional,” “lantency”), and long compound sentences could be simplified to improve clarity. The introduction would also benefit from a short paragraph connecting the presented work to recent advances in photonic tensor processors and AI accelerators, which would situate the contribution within the rapidly growing landscape of photonic–machine learning convergence.

We’re sorry for those careless mistakes and we have gone through the manuscript twice to make sure those are properly corrected.

Minor comments:

- Clarify whether the reported 10.58 nJ/pixel total energy includes both optical and electronic contributions or refers exclusively to the optical computation segment.

We have updated the Supplementary Information with the compression and reconstruction energy consumption data that accounts for optical and electronic components to ensure rigor:

For completeness, we provide here the energy consumption for compression and reconstruction including the control electronics. As mentioned in Section S2, the computing board consumes 48W (including the 1.152W of PSs). Thus, the energy consumption for image compression is calculated as $1.26 * (48 + 0.06026) / (1.152 + 0.06026) = 49.95$ pJ/pixel. The energy consumption for image reconstruction is $10.55 + 0.0303 * (48 + 0.06026) / (1.152 + 0.06026) = 11.75$ pJ/pixel.

Furthermore, constraints in commercially available electronic components prevent the PCC from operating at its intrinsic 30 GHz frequency, limiting the full system to 1 GHz. Accounting for this limitation, the energy consumption for image compression becomes $49.95 \text{ pJ/pixel} \times 30 = 1.5 \text{ nJ/pixel}$. The corresponding reconstruction energy is calculated as $10.55 + 1.20 \times 30 = 46.55 \text{ pJ/pixel}$.

This still represents an order-of-magnitude improvement in compression energy and two orders of magnitude in reconstruction energy compared to conventional electronic approaches.

- Provide a brief discussion on the reconfiguration speed and programmability of the phase shifters, commenting on the trade-offs between thermo-optic tuning (high power, slow response) and potential alternatives such as carrier-injection or microheater-free architectures.

The phase shifters are primarily employed for configuring the weight matrix of the PPC. When modifying the matrix weights, the system requires only 100 μs to reach full stabilization. We fully acknowledge the inherent limitations of thermo-optic tuning, particularly its high power consumption and relatively slow response. However, within the complete computing board, its power consumption constitutes only a minor portion, and its response speed already meets our system requirements. Adopting carrier-injection tuning would introduce nonlinear response characteristics, complicating our device-level simulations. Furthermore, this approach would lead to higher optical losses, necessitating compensation at the detection stage that would ultimately counteract any potential power savings.

- Include uncertainty estimates or error bars for key quantitative parameters such as latency, energy per MAC, and PSNR to substantiate the numerical claims and convey experimental repeatability.

These results are shown in reviewer's 1 b and 3 b comments above. Specifically, the latency per MAC is solely related to the bandwidth and decreases as the bandwidth increases, exhibiting an inverse proportional relationship. Energy per MAC is related to both bandwidth and power, showing a positive correlation with power and an inverse proportional relationship with bandwidth. As the compression ratio increases, metrics such as PSNR continue to decrease.

- Correct remaining typographical errors and ensure consistent notation throughout the manuscript (e.g., uniform use of subscripts in PSNR equations, correct formatting of units, consistent use of " $\times$ " versus "x" for dimensional notation).

We're sorry for those careless mistakes and we have gone through the manuscript twice to make sure those are properly corrected.

- Consider adding a concise summary table in the Supplementary Information comparing the theoretical and experimental parameters of the OECP (modulator bandwidth, PS efficiency, photodetector responsivity, FPGA timing, etc.) for clarity and reproducibility.

We have included a table in the Supplementary Information summarizing the performance of the optical devices. Since all our device parameters are based on measured data, no theoretical

values are provided. The performance parameters of other devices are also listed in Supplementary Information S2.

Component	Key Parameter	Value
Waveguide	Propagation Loss	1.80 dB/cm
Waveguide Crossing	Insertion Loss	0.029 dB
MZI	Insertion Loss	0.1dB
Grating Coupler	Insertion Loss	4dB
	3 dB Bandwidth	60nm
	1 dB Bandwidth	35nm
Photodetector	Responsitivity	0.95A/W
Modulator	Efficiency ($V\pi\cdot L\pi$)	2 V·cm
	doping area	1.5mm

Table R3 Key Parameters and their values of optical components.

Reviewer #1 (Remarks to the Author):

The authors' responses fall short of addressing the concerns raised. Many key issues were not addressed directly, and several critical questions remain unanswered. The rebuttal does not resolve the fundamental doubts regarding the validity and credibility of the work. Given the lack of clear novelty and the insufficient methodological rigor, this work is not suitable for publication in Nature Communications.

We appreciate the reviewer's feedback and take this opportunity to further clarify the points raised. In the first round of review, we addressed 11 comments from the reviewer, and 8 of these received no follow-up queries, indicating that the initial responses were acceptable. The 5 comments in this round are follow-ups to certain comments raised in the first round, focusing primarily on performance metrics (e.g., power consumption of electronic components and comparisons with JPEG), which we will elaborate on in detail below. We hope our supplementary analyses and revisions will fully address the reviewer's concerns and demonstrate the scientific rigor and novelty of our work.

1. The proposed work essentially applies on-chip photonic interferometric computing to the task of image compression and reconstruction. The underlying methodology follows approaches that are already well established in the field. Moreover, the proposed system still relies on electronic computation to implement FC layers, residual connections, yet these electronic components are selectively excluded from the performance and energy evaluations, which is confusing and undermines the rigor of the analysis.

This comment aligns with Comments 1 and 6 in the first round of review. We note that the reviewer's conclusion may stem from incomplete access to the supplementary materials of our revised manuscript. In the Supplementary Information (SI) of the first revised version, we have already presented a comprehensive analysis of system-level power consumption, including all electronic components (FPGA, ADC, DAC, TIA, etc.). Based on this holistic accounting, we evaluated system latency and energy consumption, and the results still demonstrate an order-of-magnitude advantage over conventional electronic approaches. Reviewers 3 and 4 raised similar concerns but confirmed that this issue was satisfactorily addressed after reviewing our updated SI.

To reinforce this point, we quote the relevant content from Sections S2 and S3 of the SI:

S2: The whole computing board integrates the PCC and all necessary control electronics (TIA,

ADC/DAC, FPGA etc.). It operates at a rated voltage of 12V and a current of 4A, with a total power consumption of 48W. The power consumption of the laser source is 17.8 dBm.

S3: For completeness, we provide here the energy consumption for compression and reconstruction including the control electronics. As mentioned in Section S2, the computing board consumes 48W (including the 1.152W of PSs). Thus, the energy consumption for

image compression is calculated as $1.26 \times (48 + 0.06026) / (1.152 + 0.06026) = 49.95$ pJ/pixel. The energy consumption for image reconstruction is $10.55 + 0.0303 \times (48 + 0.06026) / (1.152 + 0.06026) = 11.75$ nJ/pixel.

Furthermore, constraints in commercially available electronic components prevent the PCC from operating at its intrinsic 30 GHz frequency, limiting the full system to 1 GHz. Accounting for this limitation, the energy consumption for image compression at full capacity becomes $49.95 \text{ pJ/pixel} \times 30 = 1.5 \text{ nJ/pixel}$. The corresponding reconstruction energy at full capacity is calculated as $10.55 + 1.20 \times 30 = 46.55 \text{ nJ/pixel}$.

2. In the response to the comparisons with JPEG, the authors argue that the proposed work is fundamentally different from conventional compression methods. I respectfully disagree with this viewpoint. The authors are expected to comprehensively evaluate the proposed system. Moreover, the authors themselves cite figures from Ref. [1] demonstrating that photonic image compression can achieve performance comparable to JPEG in certain scenarios. This evidence, in fact, underscores the necessity of including comparisons with JPEG and other classical algorithms.

We respect the reviewer's perspective and have supplemented experiments comparing our system with JPEG using 11 test images. The resulting curves of PSNR, SSIM, and MS-SSIM are presented below (Figs. R1 – R3). Consistent with Ref. [1], JPEG achieves superior reconstruction quality at low compression ratios. However, our photonic system offers two critical advantages: significantly higher adjustable compression ratios (up to $256\times$, vs. JPEG's typical maximum of $50\times$) and orders-of-magnitude improvements in speed and energy efficiency — key metrics for real-time, low-power applications (e.g., edge computing, autonomous systems) that JPEG cannot address.

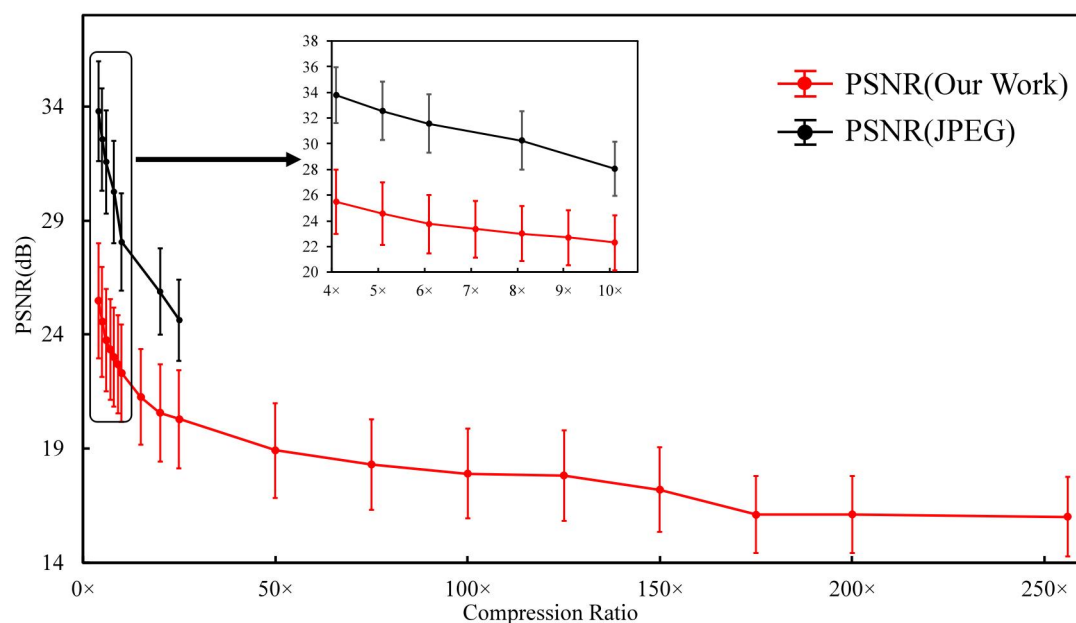


Fig. R1 Average PSNR versus Compression Ratio of 11 images.

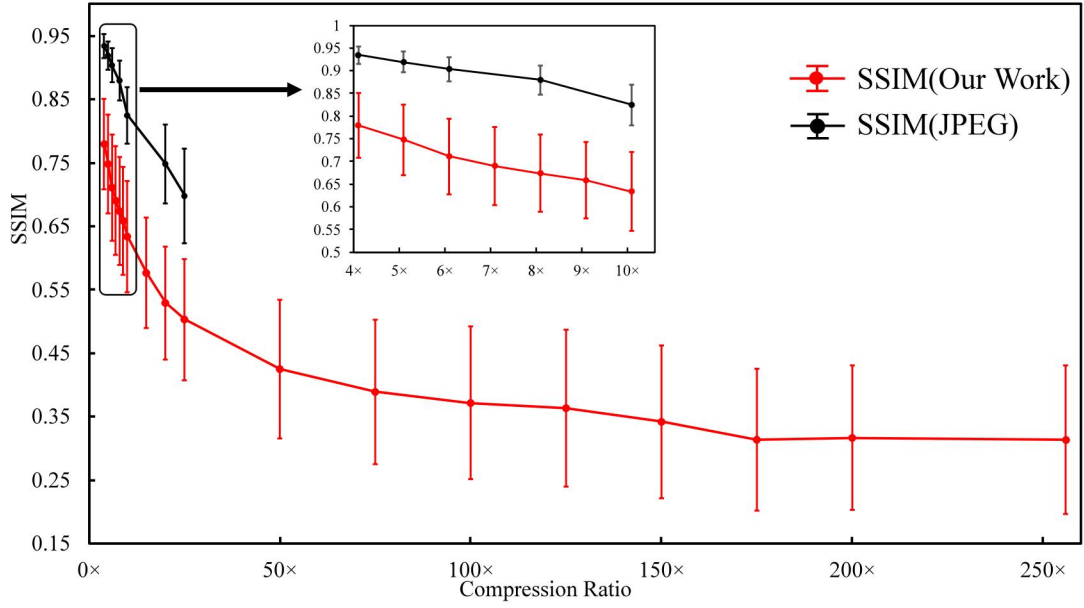


Fig. R2 Average SSIM versus Compression Ratio of 11 images.

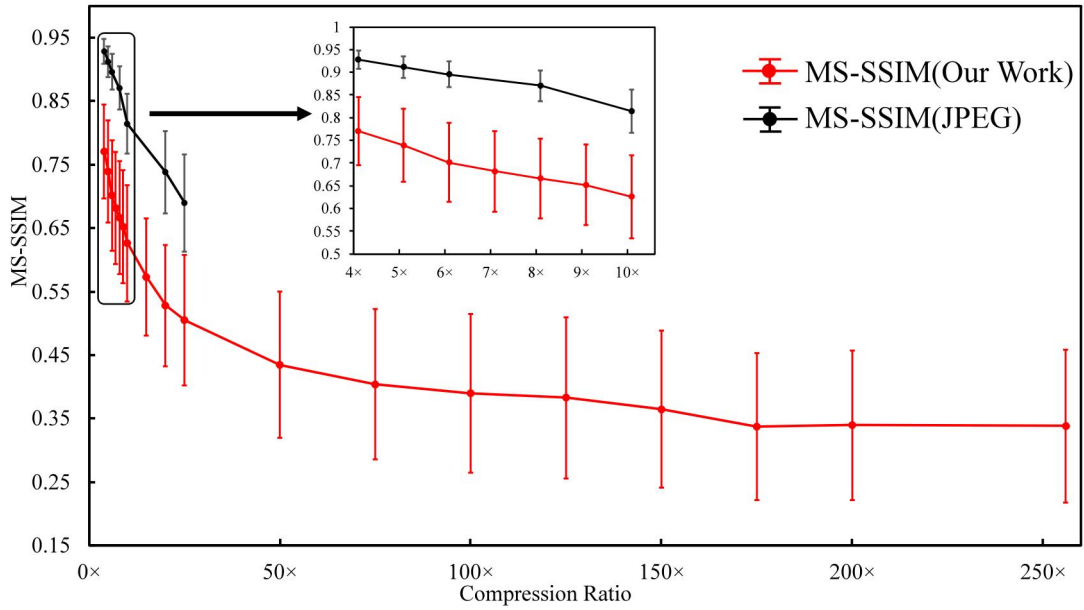


Fig. R3 Average MS-SSIM versus Compression Ratio of 11 images.

3. The authors argue that excluding AD/DA power from the energy analysis is common practice. However, even in the cited works, this exclusion is typically justified only when reporting core optical computation energy or when AD/DA occurs solely at the system output. In contrast, system-level energy evaluations generally include AD/DA and related electronic components. Ref. [2] reports modulator power consumption, Ref. [3] focuses on frequency comb power; Ref. [4] considers convolution operations only and explicitly emphasizes “considering only optical contributions”, and Ref. [1] reports solely the power of the image compression module.

The manuscript claims system-level energy efficiency, yet the proposed system requires AD/DA conversions within the processing pipeline rather than only at the output. As such, the

associated energy cannot be neglected. While the power of the compression and reconstruction modules in Table 1 may remain unchanged, the end-to-end system energy should include the overhead between these stages. The statement in the Supplementary Materials that “ADC/DAC/TIA contributes negligible energy consumption” is also potentially misleading and should be revised.

This comment overlaps with Comment 1, and we confirm that AD/DA, TIA, and other electronic components are fully included in our system-level energy calculations (as shown in SI Sections S2 and S3). To eliminate ambiguity, we have removed the misleading sentence “ADC/DAC/TIA contributes negligible energy consumption” from the SI and revised the relevant sections to explicitly state that all electronic overheads (including in-pipeline AD/DA conversions) are accounted for in the energy analysis. This ensures our system-level energy efficiency claims are accurate and comparable to state-of-the-art works.

4. The latency estimation for other related works appears to adopt calculation methods that differ from those used in the original references. The authors may need to explicitly describe the adopted latency calculation methodology in the Supplementary Materials.

To ensure fair and consistent latency comparisons, we re-evaluated the latency of related works using a unified hardware platform: a single NVIDIA RTX 4090 GPU. The codes for ReconNet, DR²-Net, and the work by Wang et al. are publicly available, allowing us to run multiple iterations and report average latency values. This approach addresses the inconsistency caused by varying hardware platforms (e.g., older GPUs) used in the original publications, enabling a direct comparison of the intrinsic performance of each model.

To further address the reviewer’s concern, we quote the original text from the related works to demonstrate the fairness of our comparison methodology.

- *RecoNet*: At the time ReconNet was published, the per-pixel latency was $0.02 \text{ s} / (256 \times 256) = 0.31 \mu\text{s}/\text{pixel}$, which is significantly higher than the $8.83 \text{ ns}/\text{pixel}$ reported in our work. This difference arises from advances in GPU performance as well as optimizations in deep learning software frameworks: For ReconNet and SDA, we use a Nvidia GTX 980 GPU to compute the reconstructions. Time complexity (in seconds) of various algorithms (without BM3D) for reconstructing a single 256×256 image. By taking only about 0.02 seconds at any given measurement rate, ReconNet can recover images from CS measurements in real-time.
- *DR²-Net*: The average latency of the Res4 model in DR²-Net is 0.0574 s; the per-pixel latency is $0.0574 \text{ s} / (256 \times 256) = 0.87 \mu\text{s}/\text{pixel}$. This value is also much higher than our measured result of $18.39 \text{ ns}/\text{pixel}$ for the same model due to advanced GPU used in our evaluation.

TABLE III
TIME (IN SECONDS) FOR RECONSTRUCTING A SINGLE 256×256 IMAGE.

Models	MR=0.01	MR=0.04	MR=0.10	MR=0.25
SDA	0.0045	0.0025	0.0029	0.0042
ReconNet	0.0326	0.0301	0.0309	0.0461
fc_{1089}	0.0028	0.0032	0.0037	0.0041
$fc_{1089-Res_1}$	0.0190	0.0200	0.0198	0.0185
$fc_{1089-Res_2}$	0.0344	0.0315	0.0349	0.0365
$fc_{1089-Res_3}$	0.0508	0.0480	0.0433	0.0505
$fc_{1089-Res_4}$	0.0600	0.0576	0.0565	0.0557

Fig. 1 Reconstruction latency of DR²-Net

- *Ref.[1]*: Wang et al. did not report the reconstruction latency explicitly; therefore, by running the model provided by the authors, we obtained a reconstruction latency of 61.67 ns/pixel, which is the highest compared with ReconNet and DR²-Net.

5. The system energy is reported as 10.58 nJ/pixel in the Introduction and energy analysis, but as 10.58 pJ/pixel in the Discussion. If this is not a typographical error, the authors should clarify the discrepancy and provide the relevant calculations.

Related original text: “Heterogeneous integration of all necessary control electronics establishes a unified photonic-electronic pipeline delivering unprecedented latency (49.5 ps/pixel) at 10.58 pJ/pixel.”

We apologize for this typographical error. The correct system energy consumption is 10.58 nJ/pixel, and we have corrected this discrepancy throughout the manuscript (including the Discussion section). The original SI calculations (Sections S2 and S3) remain accurate, and the corrected value aligns with these analyses.

The performance metrics in the present form are evaluated in an unfair manner, and the claimed innovation does not demonstrate clear advantage over Ref. [1] and related work. The manuscript does not meet the publication standards of Nature Communications.

We apologize for any lack of clarity regarding our performance metrics and innovations. To address this, we have revised the manuscript to include full electronic power contributions, removed misleading statements, and supplemented comparative experiments (e.g., JPEG). Below, we emphasize four distinct innovations that distinguish our work from Ref. [1] (the closest related work) and establish clear advantages:

1. **Adjustable and much higher compression ratio:** Our system supports compression ratios from $2\times$ to $256\times$, while Ref. [1] is limited to a fixed $4\times$ ratio. This flexibility enables adaptation to diverse application scenarios (e.g., low-latency edge computing vs. high-fidelity medical imaging).
2. **End-to-end image compression and reconstruction on photonics platform:** This is the first work to demonstrate both compression and reconstruction via photonic computing. Ref. [1] only implements optical compression, relying on external electronics for reconstruction—

undermining the speed and energy benefits of photonic architectures. Our end-to-end pipeline maximizes photonic computing advantages (low latency, high bandwidth, low power).

3. High level of integration: We monolithically integrate modulators, detectors, and a 32×32 programmable photonic matrix network, co-packaged with ADC/DAC, TIA, and FPGA. This enables standalone operation without external hardware. In contrast, Ref. [1] uses a passive photonic device with off-chip modulation and detection (despite misleading on-chip depictions in Fig. 1 of Ref. [1]), limiting integration and scalability.

4. Much lower optical loss and parasitics: Our system uses pure waveguide elements, avoiding additional loss mechanisms. Ref. [1] employs randomly etched holes along light propagation paths, introducing significant optical loss, out-of-plane scattering, and back-reflections—critical bottlenecks for integrating active components (e.g., laser cavities) and scaling to larger systems.

These innovations, combined with our revised performance metrics (fully inclusive of electronic components), demonstrate clear advantages over Ref. [1] and other related works, supporting the suitability of our manuscript for Nature Communications

Reference:

[1] X. Wang, B. Redding, N. Karl, C. Long, Z. Zhu, J. Skowronek, S. Pang, D. Brady, and R. Sarma, "Integrated photonic encoder for low power and high-speed image processing," *Nat. Commun.* 15, 4510 (2024).

[2] Y. Shen, N. C. Harris, S. Skirlo, M. Prabhu, T. Baehr-Jones, M. Hochberg, X. Sun, S. Zhao, H. Larochelle, D. Englund, and M. Soljačić, "Deep learning with coherent nanophotonic circuits," *Nat. Photonics* 11, 441–446 (2017).

[3] X. Xu, M. Tan, B. Corcoran, J. Wu, A. Boes, T. G. Nguyen, S. T. Chu, B. E. Little, D. G. Hicks, R. Morandotti, A. Mitchell, and D. J. Moss, "11 TOPS photonic convolutional accelerator for optical neural networks," *Nature* 589, 44–51 (2021).

[4] J. Feldmann, N. Youngblood, M. Karpov, H. Gehring, X. Li, M. Stappers, M. Le Gallo, X. Fu, A. Lukashchuk, A. S. Raja, J. Liu, C. D. Wright, A. Sebastian, T. J. Kippenberg, W. H. P. Pernice, and H. Bhaskaran, "Parallel convolutional processing using an integrated photonic tensor core," *Nature* 589, 52–58 (2021).