I thank the Reviewers for reading this manuscript and providing helpful feedback. Below are responses to the individual comments. Throughout this document, text added to the manuscript in response to the reviewer comments is in green.

Reviewer 1 Comment 1: The paper does not describe the configuration of the system. I think the block diagram of the system the author is developing should be clearly shown in section 2. The role of each block and the devices used in each block should be briefly explained. In this version, the analogy of the optoelectronic system and human information processing system is emphasized. Though this is an interesting analogy, the readers can not follow the detail of the operation of the optoelectronic hardware.

Response: I agree with the reviewer. I have added Fig. 3 to address this important point. The following text was also added to Sec. 2 to explain the figure:

"Schematic illustration of the neurons and modular networks under consideration are illustrated in Fig. 3. A neuron with a complex dendritic tree is shown in Fig. 3(a). Neurons with excitatory (S_e) and inhibitory (S_i) synapses feeding into dendrites (D) and the neuron cell body (N). Upon reaching threshold, the transmitter (T) produces pulses of light that fan out across a network of waveguides (not shown). Modular hierarchical construction is depicted in Fig. 3(b). The smallest blocks represent neurons, and their connections predominantly reside within their local module (blue). Yet important connections are made at all levels of hierarchy (red and dark green)."

Reviewer 1 Comment 2: The bottleneck of the optoelectronic artificial intelligence hardware is not clear. Which part limits the system performance?

Response: This is a valid and valuable question. I have responded with the following text in the discussion:

"The approach to optoelectronic hardware described here is not without limits, and different factors limit performance at different scales. Regarding speed, the synaptic response is limited by the reset time of the SPD, which is between 10\,ns and 50\,ns depending on the material used. A response time of 50\,ns limits the maximum neuronal firing frequency of the neuron to 20\,MHz. For the silicon light sources we have primarily been pursuing, the emitter lifetime is on the order of 40\,ns \cite{buta2020}, giving a maximum firing frequency comparable to the 20\,MHz figure determined by the speed of SPDs. In biological neural systems, conduction delays are an important factor limiting speed. Using light for communication greatly alleviates this concern, yet there does exist a scale at which the speed of light becomes the limiting factor. Within the 50\,ns reset time of the SPD or the comparable 40\,ns lifetime of the silicon emitters, light can travel 10\,m in fiber. A system of this linear extent would contain at least an order of magnitude more neurons than an entire human brain. The scale set by this speed limit does not represent the maximum possible scale of an optoelectronic neural system, but rather the maximum possible volume of neurons that can communicate within the highest frequency oscillations of the system.

"Regarding power consumption, cryogenic cooling plays a key role. The power required for cooling contains two contributions: the base-level power required to keep the environment below the superconducting transition temperature, even when the devices are inactive, and the additional cooling power required to remove excess heat generated by the activity of the circuits. The first factor is a few hundred watts for small systems, while the second factor is typically about one kilowatt of extra cooling power per watt of power dissipated by the devices. For small systems comprising a few thousand neurons each with a few hundred synapses on a 1cm x 1cm die, the devices will dissipate around a milliwatt \cite{sh2019}, so the first factor dwarfs the second. The second factor does match the first until intermediate-scale systems with tens to hundreds of interconnected wafers, each dissipating 1W when active. It is somewhere between the scale of a few thousand neurons on a die and a few million neurons interconnected across several wafers that we expect the performance of the system to exceed what can be accomplished without photonic communication and superconducting electronic computation. For large systems in excess of hundreds of interconnected wafers, the power dissipated by the active devices on the wafer and the associated cooling costs dominate. The power consumed by each wafer contains contributions from light sources, detectors at synapses, and JJs performing computations within dendrites and neurons. If light sources can be realized with 1\% efficiency, each of these circuit components will contribute nearly equally to the total system power consumption \cite{sh2020}."

Reviewer 1 Comment 3: The power consumption and power efficiency of the system also should be discussed. Is the use of the hardware described in this paper advantageous in terms of power consumption? If it has advantages, how large advantages of including the cooling cost of the superconducting circuits? Because the future large-scale system is large, the cooling cost is thought to be large. The power consumption of LEDs also looks large.

Response: Yes, again, good questions. These have all been answered in the text added in response to the comment directly above.

Reviewer 2 Comment 1: In the abstract the term 'general intelligence' could be very briefly introduced. In the same abstract, the 'white matter' is abruptly introduced without any explanation.

Response: The term "general intelligence" has been defined in the first sentence of the abstract: "General intelligence involves the integration of many sources of information into a coherent, adaptive model of the world." The term "white matter" has been removed from the abstract and replaced with the term "tracts".

Reviewer 2 Comment 2: In the introduction, the relation could be clarified between Ref. [11] and the following sentence, where on-chip electronic networks are discussed.

Response: I thank the reviewer for raising this point, but I respectfully decline to implement this suggestion. Taking time and space to clarify this relation will detract from the main thrust of the introduction, while adding little to the main subject of the paper.

Reviewer 2 Comment 3: In the same introduction, 'distributed optoelectronic spiking neurons' could be better contextualized.

Response: I have changed the word “distributed” to be “networked”. I have further clarified what I mean by “optoelectronic spiking neurons with the following text, added to the third paragraph of the introduction: “Spiking neurons are circuits that integrate signals over time and produce pulses when a threshold is reached. The spiking neurons discussed here are optoelectronic in that the pulses communicated from neurons to synapses consist of photons, while the computations performed within the neurons utilize electronics. Each neuron contains a light source, which is driven electrically upon reaching threshold. Each synapse contains a detector, which converts the optical signal to an electrical current or voltage upon receiving a photonic synapse event. Each neuron is a separate entity, and no hardware components are multiplexed to represent the operations of separate neurons at different times.”

Reviewer 2 Comment 4: In Fig. 1, the acronym SOEN is used before its definition.

Response: This portion of the figure now reads, “Column of 300-mm optoelectronic wafers”.

Reviewer 2 Comment 5: In Sec. 2, the fact that theta oscillation time scale cannot be exceeded could be detailed.

Response: This comment has been addressed in conjunction with Reviewer 3 comment 13. I have reorganized Sec. 2 so the definition of theta oscillations and a description of their function occurs before the discussion Reviewer 2 refers to here. I have also added additional details regarding the operation of neural systems that addresses this point. The text reads:

“In the temporal domain oscillations and synchronization structure the activity of populations of neurons \cite{bu2006}. The spiking activity of neurons is observed to comprise nested oscillations across a range of frequencies \cite{budr2004}. On the fastest time scales of the brain, local clusters of neurons engage in transient dynamical activity induced by the present stimulus. These patterns of activity are referred to as gamma oscillations (80\,Hz), and activity in this band is modulated by lower frequency oscillations \cite{caed2006,jeco2007} resulting from the combined activity of neurons across larger regions of the network \cite{stsa2000}. These slower, broader patterns are referred to as theta oscillations (6\,Hz), and neuronal communication across a network depends upon information present in gamma activity being structured into more complex syntax by dynamics on theta timescales \cite{fr2015,bu2019}. This rich structuring of information in time is enabled by the spiking behavior of neurons. Computation and communication based on spikes facilitate a diversity of information coding schemes with resilience to noise while maintaining high energy efficiency due to sparse activity.

“In the spatial domain a feature of neural systems that will recur in the present discussion is their modular, hierarchical construction \cite{sp2010,mela2010,beba2017,khma2018}. Neural systems are modular in that they are comprised of local regions of densely interconnected structures with sparser connectivity between such regions. Neural systems are hierarchical in that this pattern repeats across spatial scales in a fractal manner”

"Communication between distant modules is enabled by power-law scaling: the number of connections being sent to distant modules does not decay exponentially, but rather follows a power law \cite{bagr2010,spte2016}. The non-vanishing tail of long-range connections enables distant modules to quickly become correlated. In constructing hardware for artificial intelligence, it is imperative to enable rapid communication without traffic-dependent bottlenecks. Modules must be able to quickly engage in gamma activity, while signals from many interconnected modules at multiple levels of hierarchy must be able to simultaneously transmit across the complex network. The specific time scales defining behavior analogous to gamma and theta oscillations will be determined by the underlying computational devices."

To make matters more clear, I have also renamed Sec. 2 from "Neural devices and architecture" to "Neuroscience as a guide".

Reviewer 2 Comment 6: In Sec. 2, the Josephson junction synaptic circuit could be briefly recalled, as it stands out as the core of the proposed architecture.

Response: Such a figure has been added (Fig. 4), and has been made as compact as possible. Addition of this figure required slightly more explanation, now present in Sec. 3:

"The signals from many synapses can be combined through transformers coupled to dendrites or neurons (Fig.\,\ref{fig:synapse}(b)). Neurons constructed in this manner are highly modular in that synapses, dendrites, and the neuron cell body itself are all based on the same core circuit, comprising a superconducting quantum interference device (SQUID) embedded in a flux-storage loop. SQUIDs are perhaps the most ubiquitous of all superconducting circuits \cite{vatu1998,ka1999}, often used as sensors due to their extraordinary sensitivity to magnetic flux and low-noise operation. These properties make SQUIDs ideal circuits for dendrites and neurons to perceive and respond to minute changes in analog signal levels."

This text has also been added to the same paragraph for clarity:

"and can be shaped through the choice of circuit parameters, such as loop inductances and resistances, as well as dynamically with adaptive bias currents."

Reviewer 2 Comment 7: Fig 2(b) could be detached and placed where it is described. On the other hand, a graphical representation of small-world networks could be useful.

Response: I have moved this figure, as recommended. I do not wish to add a graphical representation of small-world networks, as this paper is already excessive in length and content for the purpose of a perspective article. The relevant literature containing a graphical representation of small-world networks has been cite (Ref. 27).

Reviewer 2 Comment 8: In Sec. 3, the optical link efficiency could be briefly introduced and the ways to overcome the 10^4 efficiency gap could be discussed.

Response: The following text has been added to Sec. 3 to speak to this comment:

"An important milestone was the demonstration of an all-silicon monolithic optical link."

"Subsequent work improved the brightness of the sources by two orders of magnitude through optimized fabrication procedures \cite{buta2020}. Additional gains may result from optimization of the diode structure used for electrical injection of carriers into the waveguide where electron-hole recombination at emissive centers produces waveguide-coupled luminescence. Elimination of etched surfaces and proper passivation in the active region may significantly reduce non-radiative recombination. Improvements to the optical structure may increase coupling efficiency from the emitters to the waveguide mode."

The last sentence of this paragraph was changed. It originally read, "If silicon light sources can meet these performance specifications, the hardware stands a chance of enabling brain-scale systems with 30,000 times the speed." It now reads, "Modest advances could enable silicon light source technology to meet these specifications."

To partially mitigate the extra length incurred due to the addition of these explanations, I have reduced the descriptions in the captions to Figs. 2 and 5.

Reviewer 2 Comment 9: The review of the literature in Sec. 3 about photonic neural systems would be enhanced by providing a more detailed quantitative comparison among the different approaches (again possibly summarized in a table for ease of reading).

Response: This comment has been addressed in conjunction with Reviewer 3 Comment 1. To begin, I placed the existing text reviewing the literature in a new section, entitled "The landscape of research in photonic and superconducting neural systems". Subsections have also been introduced for clarity. In this section, I have added the following text:

"It is clarifying to acknowledge that nearly all the efforts to use photonics or superconducting electronics for neural systems are focused on the entirely reasonable goal of doing useful computations with hardware that is available right now. These efforts are valuable and promising for their own ends without seeking brain-scale cognition. The superconducting optoelectronic hardware discussed here is in an early stage of development, as described in Sec. 3. Comments here contrasting SOEN hardware with other current efforts are not criticisms of any work in the field, but rather as explanation of the reasoning behind SOENs for large cognitive systems. Here I provide a short summary of other photonic and electronic efforts. For comprehensive reviews of efforts in emerging neural hardware, the reader is directed to recent reviews."

The subsection "Semiconductor electronic neural systems" was moved here from Sec. 5, as it now makes more sense to be in this broader conversation about other approaches. Following that subsection, the following text was added to the "Optical neural systems" subsection:

"One means to alleviate communication limitations is through the use of optics. The field of photonic neural systems began \cite{psfa1985,faps1985} with an implementation of the Hopfield model \cite{ho1982}. The objective was to combine the parallelism and interconnectability of optics, which are linear phenomena, with bistable optical devices to provide the thresholding nonlinearity of the Hopfield model. The hardware proposed combined compound-semiconductor LEDs with photodiodes and electronics for an initial implementation of nonlinearity to be replaced by optical bistable devices in subsequent generations. While LEDs and laser diodes have become mature technologies, bistable optical devices have not.

“The field of photonic neural systems has since experienced an immense diversification, with myriad efforts using free-space optics \cite{habe2019}, fiber components \cite{prsh2017}, and on-chip integrated photonics \cite{shha2016,vame2014,tafe2019}. Along one branch of this tree”

“Another exciting and related application space of photonics is in reservoir computing. This field has been innovative and productive in recent years \cite{funa1993,vada2011,orso2015,vabr2017,brpe2018}. The objectives and hardware are only loosely related to the subject of large-scale cognition considered here, so further discussion is omitted.”

Reviewer 2 Comment 10: The one thousand power penalty for cooling could be better explained.

Response: The text added to the discussion in response to Reviewer 1 Comment 2 addresses this comment as well.

Reviewer 2 Comment 11: In Sec. 3, the motivation behind the need of recurrent networks could be highlighted.

Response: The motivation for recurrent networks has now been further elucidate with the discussion added to Sec. 2 in response to Reviewer 2 Comment 5.

Reviewer 2 Comment 12: In Sec. 4 some more detailed considerations on the strategies for the packaging and the coupling of many fibers to photonic integrated circuits could be added, to provide a robust case for the scalability of the proposed approach.

Response: The following text has been added in response to this comment:

“Recent progress in low-loss fiber-to-waveguide coupling \cite{khbu2020} indicates a potential future direction for such integration of fibers with on-chip waveguides, but significant advances in manufacturing are required to realize the coupling of dense fiber bundles to 300-mm wafers.”

Reviewer 3 Comment 1: I find that the introduction does not provide a general overview of the field of opto-electronic neural networks - which ultimately is the title of this work. Opto-electronic NNs is a research field spanning decades of activity, and some seminal experiments or review papers not only limited to the precise field or technology platform the author is advocating would help interested readers.

Response: I agree that this is important, and this point has been driven home by the fact that multiple reviewers mentioned a similar criticism. I have given this comment significant additional attention in the context of Reviewer 2 Comment 9. Please see the response there.

Reviewer 3 Comment 2: The sheer physical scale of the discussed system. The author targets GAI and uses biological analogies to motivate the network-topological scale that would need to be targeted. It, however, remains rather unclear to the reader, besides some very punctual notes within the manuscript, that the targeted system will span meters or even 100 m length scales. A proper, visible and open discussion early in the text of this is enormous expand is fundamentally important for the read due to various reasons:

1. Such an infrastructure certainly is only interesting for a Dr. Manhattan kind of intelligence; for more every-day intelligence the associated efforts would most likely overcome the benefits. As a consequence, this manuscript should be most interesting and important for large-scale infrastructure projects, and a reader should be aware from an early stage.
2. The author leverages arguments stemming from very large-scale integration at various points of the manuscript in a 'various lessons learned from previous VLSI' kind of style. The gigantic physical scale of the suggested system is therefore very relevant on a fundamental argument scale. Many previous VLSI insights could potentially not simply be transferred to the mere physical scale envisioned.
3. Size (volume) is highly relevant for cryogenic aspects, and as the author targets heavy cryogenic usage in an integrated circuits setting it would mean that a significant fraction of this machine would have to operate at 4K. Such a GAI computing center would potentially dwarf the cryogenic volumes of other mega-structures like ITER and CERN, and for many conclusions made in this article to be founded scientifically this discussion should be at least be sketched out in first principle and quantitatively included in the energy consumption arguments .

Response: This is a good point. I want the reader to be clear about the scale of systems under consideration at the outset. I have added this paragraph at the end of the introduction:

“The unique cognitive capabilities of humans derive in part from the scale of the brain, including the number of neurons and the complexity of the communication network. While there is much to be gained from AI hardware at smaller scales, this article considers technological pathways to large cognitive systems, with tens to hundreds of billions of neurons, and communication infrastructure of commensurate complexity. Such technology will likely require many interconnected wafers, each packed densely with integrated circuits. We may refer to this field of research as ``neuromorphic supercomputing''. The effort is in some ways more akin to the construction of a fusion reactor or particle accelerator than a microchip, and potentially offering a similar scale of societal benefit in the form of an experimental test bed enabling the elucidation of the mechanisms of cognition and the exploration of the physical limits of intelligence.”

Reviewer 3 Comment 3: The manuscript makes a general and good introduction and motivation of optical communication. However, then it essentially heavily focuses on very narrow and specific superconducting technology targeting very high level of artificial intelligence. I am therefore not entirely convinced the title reflects the true focus and perspective of the manuscript, and find that it over-reaches to a certain degree. It appears to imply that electro-optical intelligence is only achievable via superconducting JJs, which is a opinion I would challenge only having the here presented arguments at hand.

Response: This is a valid criticism, which I summarize as, "The manuscript is too focused on superconductors. There could be other routes to optoelectronic intelligence." I agree, and I have made significant additions to address this, perhaps the most important of all reviewer concerns. This comment is very much related to the comment immediately below, so I address them both there.

Reviewer 3 Comment 4: The superconducting argument is motivated on a very quick and a bit superficial scale, for example: "If cryogenic operation enables both single-photon detectors and silicon light sources, it will be worth the added infrastructure for cooling." The author's argument can therefore be boiled down to "single photon detection and silicon laser sources require 4K", which I think is not correct. Avalanche photo-diodes facilitate single photon detection which does not require 4K, and laser sources based on III-V quantum dots integrated directly in the silicon platform even become commercially available. I find the "room-temperature photodetectors require several thousand photons to register an event" not correct, or certainly in the light that 4K for sure not is required.

Response: Another excellent point from the reviewer, very much related to the comment above. This point is important enough and central enough to the concept of the manuscript that I have given it considerable additional attention in the manuscript. I have added the following text to the end of section 3:

"In addition to contrasting this approach to other existing work in the field, it is necessary to also consider what may seem a more straightforward route to optoelectronic intelligence. This route would involve spiking neurons based on waveguide-integrated light sources, as we have discussed, but instead of SPDs and JJs, semiconductor photodiodes and transistors would be employed. Pursuit of such hardware is impeded by the absence of light sources integrated with transistors. If there were a known means to integrate light sources as simple as transistors with silicon microelectronics, the landscape of computing would differ radically. Nevertheless, the proposition that superconducting electronics are more promising than photodiodes and MOSFETs for this application requires justification.

"Our choice to focus on the superconducting approach is based primarily on three factors. First, superconducting single-photon detectors dramatically reduce the brightness required of the light sources. While semiconducting detectors, such as avalanche photodiodes, can detect a single photon, the energy consumption negates the benefits of single-photon sensitivity in the system application under consideration. For scalable system integration, the semiconductor counterpart to a waveguide-integrated SPD working in conjunction with a JJ is a waveguide-integrated photodiode working in conjunction with a MOSFET. Such a semiconductor receiver is likely to require roughly 1000 photons to charge the capacitance of the MOSFET gate \cite{mi2017} to initiate a synapse event. This factor of 1000 in photon power is matched by the factor of 1000 incurred to cool the superconducting system (see Secs.\,\ref{sec:communication} and \ref{sec:discussion}), so the net power consumption for light generation in semiconductor and superconductor systems is roughly equivalent. Yet the important distinction is that the superconducting system dissipates this power off chip in a cryocooler, whereas the semiconducting system requires the light sources to produce this power in the form of photons. Optoelectronic neural systems leveraging superconductors can make due with light sources providing 10,000 photons within a few tens of nanoseconds (30\,nW continuous-wave equivalent), while a semiconducting counterpart will require light sources 1000 times brighter to attain the same firing rate. Achieving the former appears possible with inexpensive silicon light sources, while the latter is likely to require further advances in III-V sources. While exciting progress continues to be made in III-V integration on silicon \cite{tapa2019,hala2020}, a central challenge remains to integrate these light sources intimately with electronics. The system under consideration requires fabrication of light sources by the millions across 300-mm wafers, which will surely be more cost effective if silicon devices as simple as transistors can be employed for light emission \cite{buch2017}, a possibility that appears more likely with superconducting detectors and low-temperature operation.

“The second factor driving our group to pursue the superconducting approach relates to multi-planar wafer-scale integration. Whether semiconductors or superconductors are used, artificial synaptic, dendritic, and neuronal circuits are not small. To accommodate millions of neurons and their synapses on a 300-mm wafer, on the order of 20 planes of photonic waveguides are required for communication, and a similar number of planes of electronic circuits are likely to be advantageous. For each plane of MOSFETs, high-temperature annealing steps are required for dopant activation, leading to processing challenges when integrating with metal wires, photonic waveguides, and light sources. This processing challenge is one reason extension of MOSFET processes to multiple stacked planes of transistors with copper interconnects has been difficult. Power dissipation and heat removal also come into play but may be less consequential in the context of spiking neurons with sparse activity. Superconducting electronic circuits are processed near room temperature, and the prospect of integrating many planes of JJs, SPDs, and waveguides appears to us to be less restrictive. Multiple planes of active SPDs\cite{vema2012} and JJs \cite{tobo2019} have been demonstrated.

“The third factor steering us toward superconducting electronics relates to memory and learning. For a cognitive system of the scale under consideration, synaptic weight modification must be unsupervised and will be most readily realized if the signals that induce learning functions are the same signals, with the same current, voltage, or light levels, used for computing within neurons, and sent to synapses for communication. With superconducting circuits, single-flux quanta are used for computing, and single photons are used for communication. It appears possible for these same signals to update synaptic weights and enable learning, primarily by adjusting current biases to JJs. A close functional analogy would be to modify the voltage on the gate of a MOSFET in an analog manner, and indeed, this has long been the ambition of floating-gate MOSFETs for synaptic memory \cite{hama2013}. However, the voltages required to change the charge on the gate are much higher than typical voltages used for computation elsewhere within the circuit, making it difficult to implement unsupervised learning based only on the signals already present in the network. These persistent challenges with floating gates have led many to look elsewhere for suitable adaptive circuits \cite{upji2019}. While any one of these approaches may lead to the desired memory operations, it is our perspective that the path to systems with lifelong learning and a multitude of memory mechanisms appear less formidable with Josephson circuits.”

Additionally, I have changed the title of Sec. 3 from “Optoelectronic synapses, dendrites, and neurons” to “Superconducting optoelectronic synapses, dendrites, and neurons” to further clarify the specific technology under consideration.

Reviewer 3 Comment 5: Besides components, the cryogenic nature of the concept is mostly motivated by the low power consumption potentially accessible when operating at this single information quanta limit. This for me is a non-trivial argument at several scales and should be expanded and justified more careful.

1. Single photo detection is often an energetically very inefficient process. Based on the here leveraged superconducting wire detectors, photon absorption induces local heating that breaks superconductivity as the system is operated close to its critical current. Energy consumption I therefore assume is mostly dominated by dumping this entire breakdown energy into the substrate, and not the single photon absorption.
2. The consequence of the super-conducting part is that power-splitters required for the network connections cannot be properly implemented as only 1x2 splitters exist. The author writes "The challenge with using superconducting electronics is communication. In superconducting circuits, direct fan-out is usually limited to two." If I understand correctly then that implies that each weight is tied to a detector. While one might be able to reduce the number of required photons propagating it appears to me that on the same scale one requires more detectors. As detection is always less efficient than propagation (at least now) I would like the author to comment a bit more in detail.

Response: Again, I agree with the reviewer. I have added this text to Sec. 4:

"Despite these arguments in favor of superconducting electronics, several valid counterpoints can be raised. The requisite silicon light sources remain to be proven. Massively multiplanar fabrication of superconducting optoelectronic wafers is an ambitious technological undertaking. For many readers, the requirement of cryogenic operation is the most disconcerting aspect of the project. Several comments are in order. Low-temperature operation eliminates such systems from consideration for applications that require low system power consumption, such as mobile devices. But for systems with a million neurons, existing cryogenic technologies drawing a kilowatt of wall power are suitable, comparable to a home air conditioner in power consumption and complexity, but with cooling based on the thermodynamic properties of liquid helium. For larger applications, cryogenic operation may prove an insurmountable obstacle, although the scale of cryogenics used in superconducting magnets for particle colliders offers hope. The field of quantum information also provides an insightful lesson. Many types of qubits require operation at tens of millikelvin, necessitating the extra expense and complexity of dilution refrigerators. The environment at 4\,K is comparatively balmy, and the required cryogenics are simpler and less expensive. Quantum information presently enjoys tremendous investment because these systems promise functions not otherwise possible. The same must be true of optoelectronic intelligence if it is to have a future. Anything that can be done with CMOS will be done with CMOS. If SOENs cannot achieve AGI that is otherwise unattainable, they will not be brought into existence. If they can attain unmatched cognition, someone is likely to be willing to pay for them, unless the expense is astronomical. The perspective presented here is that exactly this will come to pass: superconducting optoelectronic hardware will enable AI that simply cannot be achieved through other physical means. Low-temperature operation will be justified by the performance."

Reviewer 3 Comment 6: Spiking networks are an attractive and interesting concept, yet the particular value for computation and more importantly for hardware integration is not as immediate as implied by the author on various positions of the manuscript. Often the argument for energy efficiency in all-or-nothing signals is leveraged, yet this ignores the cost of keep the components close to criticality. The brain spends a major part of its energy on DC potentiation. The same is true for signal to noise ratio, as this delicately depends on details of signal encoding. I am far from against taking a strong stand for spiking NNs, yet for rigor these features should be discussed or at least mentioned in order to provide a reader with a more representative notion of these concepts.

Response: The following text has been added to address this criticism. In Sec. 1, content was already added to address a similar concern raised by Reviewer 2 Comment 3. In Sec. 2, this text was also to address Reviewer 3's comment here:

"Computation and communication based on spikes facilitate a diversity of information coding schemes with resilience to noise while maintaining high energy efficiency due to sparse activity."

Reviewer 3 Comment 7: Single photon as an adequate carrier of spike-encoded information. Does operating the system at the physical limit of 'spiking' signals not come with a fundamental challenge? I would be worried that shot-noise will results in very, very poor SNR performance. Could you please comment on that?

Response: Yes, another important point. We have addressed this in Sec. 4 of the manuscript with the following text:

"As a broad point of contrast between the synapses discussed here and other systems using light for neural computing, most photonic neural systems encode information in the amplitude of optical signals received at a detector, and synaptic weights are established through modulation of the intensity of these optical signals. Whether phase modulation or direct amplitude modulation are leveraged, encoding synaptic weights in the intensity of light on a detector differs from the synaptic operations we are pursuing, where light is used for binary communication, and synaptic weights are established by electronic responses. This approach minimizes the optical power required and eliminates a source of noise. If synaptic weights are encoded in the intensity of an optical signal, noise from the light source is convoluted with the synaptic weight. With binary optical signaling the light level incident upon a synaptic detector does not influence the electronic response of the synapse, which is determined by the electronic circuits reading out the synaptic receiver. A binary response can be achieved with semiconductor receivers or superconducting circuits. In Ref.\,\onlinecite{buta2020_2} we have shown that the response of a superconducting SPD is independent of the number of photons present in an incident pulse across four orders of magnitude of input intensity. Because no information is encoded in the light level, this form of optical communication does not suffer from typical shot noise. Provided one or more photons are received by the detector, a synapse event is communicated. The Poisson distribution gives the probability that zero photons are received. With an average number of five or greater photons transmitted per synapse event, the probability of receiving zero photons is less than 1\%, a considerably lower error rate than biological synaptic transmission \cite{li1997}. We assume each neuronal light source will generate 10 photons per synaptic event to accommodate 3\,dB of propagation loss while achieving 99\% transmission success rate. All energy and power consumption estimates presented here use this value."

Reviewer 3 Comment 8: "In this regard, difficulties associated with integrated light sources are the most significant impediment to optoelectronic VLSI." I think this is mostly the challenge for classical circuits considered in VLSI. In terms of NNs and in particular for photonic NN implementation there might be new challenges arising, for example size scalability. Maybe it is worth mentioning that.

Response: These challenges have now been addressed though additional text added in response to previous comments. Cryogenic challenges were address in response to Reviewer 3 Comment 5. Fiber packaging and wafer integration were addressed in response to Reviewer 2 Comment 12. The large scale of the systems was addressed in response to Reviewer 3 Comment 2.

Reviewer 3 Comment 9: The author claims that meshes of Mach-Zehnder cannot implement recurrent connections, which is not correct, see review of Bogaerts, Nature 586 (7828), 207-216 (2020).

Response: I hate to be pedantic, but I actually did not say meshes of Mach-Zehnder interferometers *cannot* implement recurrent connections. I said they "are not conducive to establishing" recurrent networks. The further clarify, I have added the following text:

"The challenge arises because in meshes of interferometers, adjustment of one phase modifies multiple synaptic weights. While such a technique may be suitable for specific training algorithms employed in supervised learning \cite{humi2018}, it appears cumbersome for unsupervised learning in large neural systems, where local activity at each synapse updates that synaptic weight."

Reviewer 3 Comment 10: "and synaptic weights are established through attenuation of these signals." Is not correct. Numerous approaches, e.g. photonic reservoir (Bueno et al, Optica 2018) as well as in deep neural networks (Lin et al. Science 2018) use diffraction, which by first principle uses phase modulation and hence is mostly unitary. Some other work uses phase encoding for injection information into a semiconductor laser. The same is true for electro-optical memristors which can implement phase shifts and are being considered.

Response: I have clarified my intended meaning with changes to the paragraph that begins, "As a broad point of contrast…" The primary point of clarification is as follows:

"This approach minimizes the optical power required and eliminates a source of noise. If synaptic weights are encoded in the intensity of an optical signal, noise from the light source is convoluted with the synaptic weight. With binary optical signaling the light level incident upon a synaptic detector does not influence the electronic response of the synapse, which is determined by the electronic circuits reading out the synaptic receiver. A binary response can be achieved with semiconductor receivers or superconducting circuits. In Ref.\,\onlinecite{buta2020_2} we have shown that the response of a superconducting SPD is independent of the number of photons present in an incident pulse across four orders of magnitude of input intensity."

To make space for this additional discussion, the following text, which was redundant with discussion added in response to other comments by Reviewer 3, has been removed: “While operation at 4\,K brings a factor of one thousand power penalty for cooling, waveguide-integrated, room-temperature photodetectors require several thousand photons to register an event. Thus, the power penalty of cryogenic operation is compensated by the ability to detect single photons, even before the efficiency gains of cold light sources and superconducting electronics are considered.”

Reviewer 3 Comment 11: "Generating more than a millivolt with superconducting circuits is difficult, but the thin-film, micron-scale cryotron" Is this process compatible with Silicon integration?

Response: This comment has been addressed with the following text added to Sec. 3:

“Fabrication of these devices appears compatible with silicon microelectronic manufacturing, provided the high-temperature steps required for dopant activation and contact annealing required for semiconductor devices are performed prior to the deposition of superconducting thin films.”

Reviewer 3 Comment 12: Figure 2(b) is not discussed in the text until much later stages in the manuscript.

Response: This figure has been moved, as suggested here and in Reviewer 2 Comment 7.

Reviewer 3 Comment 13: The general reader might benefit of a definition of the "average path length" as well as on the relevance of "theta oscillations" in the brain.

Response: Theta oscillations have been addressed in response to Reviewer 2 Comment 5. To define the concept of average path length (as well as the clustering coefficient introduced at the same time in the manuscript) the following text has been added:

“In the language of network theory, if node \$a\$ is connected to node \$b\$, and \$b\$ is connected to \$c\$, then clustering quantifies the probability that \$a\$ will be connected to \$c\$. Path length quantifies the number of intermediate nodes that must be traversed to get from one node to another along the network connections. The average path length is determined by calculating this quantity over all pairs of nodes in the network, and taking the mean. A network with high clustering and low average path length is referred to as a ``small-world network'' \cite{wast1998}.”

Reviewer 3 Comment 14: The authors mentions brain scales from starting at 1 nm, which I would argue is at least an order of magnitude too small. Postynaptic receptors begin above 10 nm, and I am not sure that their effects and sizes are relevant for the here discussed context.

Response: I respectfully decline to modify the manuscript in response to this comment. Neurotransmitters sent across the synaptic cleft are on the order of 1nm. Regardless, the distinction between 1nm and 10nm is of little consequence to the point being made or the larger message of the paper.