We thank the Reviewers for reading this lengthy manuscript and providing helpful feedback. Below are responses to the individual comments.

Reviewer 1 Comment 1: In section 6, what are the definitions of q, and tau in the exponential?

Response: q is the power-law exponent, and tau is the exponential decay time constant. We have clarified with this statement in the text: "In Fig. 11(a) we compare $f(t) \sim t^{-q}$ to $g(t) \sim e^{-t/\tau}$ for three values of τ, taking the power-law exponent to be $q = 1$." The definitions are also given in the caption to figure 11.

Reviewer 1 Comment 2: The part related to Figure 11 (the text in the body and the figure) is not very clear. More explanation would be helpful.

Response: We have reworked the paragraph providing the motivation for and explanation of Fig. 11. Changes are marked in red. The paragraph now reads: "Figure 11(a) summarizes the motivation for employing two different time constants. Instead of retaining a memory trace of a synapse event over only a single temporal scale, as occurs in a single SI loop with exponential decay with one time constant, we would prefer a signal with a power-law decay, so that information across temporal scales can be accessed. In Fig. 11(a) we compare $f(t) \sim t^{-q}$ to $g(t) \sim e^{-t/\tau}$ for three values of τ, taking the power-law exponent to be $q=1$. This figure illustrates the principle that a power-law temporal decay represents information across multiple orders of magnitude in time, while an exponential function only has a single time constant, and therefore only represents information across roughly one order of magnitude. For example, in Fig. 11(a), the power-law decay has an appreciable signal spanning two orders of magnitude in time. By contrast, the smallest value of τ provides no information past its cutoff, and the signal from the largest value of τ is nearly constant initially across more than an order of magnitude. The middle value gives a poor representation at the start and the end. However, we can obtain a suitable approximation to the power law function by superposing a small number of exponentials [44], as shown in Fig. 11(b). Here we represent a power law with unity exponent ($q = 1$), mapping two orders of magnitude in time to two orders of magnitude in signal. Convergence is shown in the inset. The error is improved by an order of magnitude when using two exponentials instead of one, and there is little advantage to using more than three for this task."

Reviewer 1 Comment 3: The circuit and function of RQ is not very clear. Is it just a big inductive current that would start a switch? It seems that any switch would do that function. This part needs more/better explanation.

Response: We have added additional information regarding rapid query. The original paragraph is now broken into three paragraphs, which read as follows (additions in red): "The duration over which the dendrite is inhibited is controlled by τ_{ih}, and for the network to be rapidly adaptable under the influence of inhibition, this time constant will be as short as a gamma-range interspike interval. If inhibition is required over theta time scales, repeated activity on the inhibitory neuron can keep the dendrite suppressed. However, this may not be the most energy-efficient mode of operation. Given the circuits under consideration, we can utilize a mode of operation complimentary to inhibition. In this configuration, the mutual inductors and bias to the DR loop are chosen so that even with all afferent SI loops saturated, the current across J_{df} cannot exceed I_c. Only when an additional, unique synapse fires does the current exceed I_c.

We refer to these unique synapses as *rapid query* synapses, and we explain their function as follows. The role of a rapid query synapse is complimentary to the role of an inhibitory synapse. In the typical operation of a dendrite, the response of the dendrite depends on the activities of the input excitatory synapses, and the role of inhibition is to effectively cancel the excitatory inputs. In one sense, the function of a rapid query synapse is the opposite of the function of an inhibitory synapse. With rapid query, a dendrite is designed so that the sum of all excitatory synaptic responses is insufficient to evoke a dendritic response. However, the function of the rapid query synapse is to drive the dendrite right up to its threshold, and therefore any excitatory input present at the time the rapid query synapse fires evokes a dendritic response. The circuit implementing this unique synapse is identical to all synapses considered thus far (Fig. 2(a)), but the function is unique. This synapse is designed to saturate with each synapse event and to decay rapidly, providing an identical response to each synapse event and no synaptic weight variation. The action of this synapse is to allow J_{df} to quickly sample the value of I_{dr} at a given instant in time. A dendrite with one or more excitatory synapses and a single rapid query synapse behaves as if it were always under the influence of inhibition until the rapid query synapse fires, briefly releasing it from inhibition. When the rapid query synapse fires, the current generated in the DI loop provides an answer to the question, ``How much current is in the DR loop?''

As stated above, the function of a rapid query synapse is the opposite of the function of an inhibitory synapse. However, in another sense the objective is the same. A primary function of inhibition in neural systems is to dynamically adapt a given structural network into multiple functional networks. When inhibition is applied to a dendrite, the dendrite is functionally disconnected from the neuron cell body. Similarly, with rapid query, a dendrite is functionally disconnected from the neuron cell body at all times except when a rapid query synapse has fired. Rapid query operation provides another means to rapidly adapt the structural network into myriad functional networks, and rapid query is likely to be more energy efficient than inhibition when information stored in certain dendrites need not be accessed frequently. In biological neural systems, a given neuron either makes inhibitory connections or excitatory connections, but not both. This is referred to as Dale's law. It may be a consequence of physiological limitations, or it may be due to an information-processing advantage resulting from differentiating the responsibility of excitatory neurons that spread information and inhibitory neurons that adapt the functional network. We anticipate that such differentiation will be advantageous in superconducting optoelectronic networks as well, in which case neurons dedicated to inhibition will make only inhibitory synaptic connections, and neurons dedicated to rapid query will form only rapid query synapses. We refer to these neurons as rapid query neurons. The role of a rapid query neuron is to quickly cause the information stored in a collection of dendrites to be communicated from those dendrites further along the respective dendritic trees toward the neuron cell bodies, thereby rapidly functionally connecting those dendrites to the active network."

<u>Reviewer 1 Comment 4</u>: Although touched briefly at the end of the paper; but, the margin of few percent current, for example, for such a change in the function of the circuits is almost impossible in terms of manufacturability. For example, adjusting the current from 37.25uA to 39 uA in figure 6 is not

that easy for thousands or millions of circuits. What would happen if two circuits are not identical due to variability in fabrication?

Response: The reviewer raises a critical point that is very important to the success of this line of research. We have added additional discussion in two parts of the text to address fabrication variability. In the context of synapses, we have added this paragraph: "From Fig. 3, we can also gain some insight into the manufacturability of these circuits. With this technology, we aspire to achieve large-scale systems capable of advanced cognitive computing. Such systems will potentially comprise billions of synapses and 10 times as many JJs. These JJs will have a statistical distribution of critical currents due to fabrication variations. During operation, the biases delivered to the junctions will also have a statistical distribution. The data in Fig. 3(b) inform us that synapses with a broad range of bias conditions will contribute signal upon receiving synaptic events. Here we show that synapses will be operational if the bias current varies by 8 µA around 38 µA, giving a margin of 20%. When the system is initially fabricated and turned on, variations in junction I_cs and biases will result in a statistical distribution of synaptic weights. Over time, as the system operates and learns, these bias currents will be finely adjusted based on the activity-dependent plasticity mechanisms described in Ref. 13, mitigating any deleterious effects of fabrication variations." In the context of coupling a synapse to a dendrite, we have added this paragraph: "As mentioned above in the context of synapses, we wish to anticipate how fabrication imperfections in JJ critical currents as well as variations in bias conditions will affect circuit operation. The data in Fig. 6 show that the dendritic response is quantitatively sensitive to the value of the bias currents, or similarly the junction critical current. However, the qualitative nature of the response is consistent across a useful range of operating parameters. In large-scale systems, the intention is not to precisely control the response of each dendrite or synapse quantitatively at the time of fabrication, but rather to fabricate a complex network with a statistical distribution of device parameters and to employ adaptive plasticity functions that finely adjust biasing condition through activity-dependent feedback, as discussed in Ref. 13, to adapt the circuits to operating points useful for network computation. Such adaptation over time through synaptic and dendritic plasticity are in the spirit of biological neural systems that cannot be constructed with explicit values for each synaptic weight or precise dendritic morphology. Nevertheless, it remains to be seen if Josephson circuits can be manufactured with tight enough tolerances to enable the functions proposed here. This question is one of the most pertinent to determining the feasibility of large-scale superconducting optoelectronic networks."

Reviewer 1 Comment 5: The argument for making the fluxonic energy and photonic energy equal was not very clear or may need more explanation. Minimizing the footprint and energy would seem to be a goal for the design.

Response: We have clarified our thinking with the following changes: "In mature superconducting optoelectronic circuits, system-level considerations will inform decisions made regarding trade-offs between energy consumption and performance. One could reduce energy and area consumption by omitting dendritic processing entirely, but this would leave out important information processing. We do not attempt to fully address these trade-offs here, but we briefly consider the energy expended on synaptic, dendritic, and neuronal operations. At the physical level, these operations require light production during neuronal firing, photon detection during synapse events, and fluxonic processing in the dendritic tree. We would like the energy expended on light production, photon detection, and fluxonic processing to be roughly equal. Such an operating point is appealing because it indicates a global optimum wherein improvements to any one aspect of the system provide little added benefit, as other contributions become limiting factors. We can estimate where this operating point may reside by considering the three primary contributions to energy consumption. Similar analysis can be conducted regarding area."

Reviewer 1 Comment 6: Comments about the manufacturability of these ideas would be very helpful.

Response: I have added the two paragraphs discussed above in the context of Comment 4. In the interest of space, I will refrain from further discussing manufacturing in the conclusions.

Reviewer 2 Comment 1: It seems that the general premise for the purpose of designing and considering the technology and architecture described is primarily to use the set of technologies described as a platform to investigate biological neural systems. This is not as clearly put forward in the introduction as it could perhaps be. It might be helpful to differentiate "biological neurons" and neural systems from the "artificial neurons" described here - particularly in the introduction and summary.

Response: The anticipated uses of this technology are clarified at the beginning of the third paragraph. We have added this text: "The present work is concerned with artificial hardware capable of neural information processing. Such hardware is anticipated to be used both in the scientific study of the mechanisms of neural information processing as well as in technological applications that benefit from neuromorphic computing. In previous studies, we have considered artificial hardware based on superconducting optoelectronic circuits to achieve point-neuron functionality [11-13]. The present work builds on those circuit concepts, introducing new superconducting electronic circuitry to perform functions associated with dendritic processing in biological systems." We have also added the words "biological" and "artificial" to the abstract to clarify.

Reviewer 2 Comment 2: I don't believe that NR in figure 2 is explicitly defined in the text, and should be included in the caption.

Response: I have added the definition of NR to the caption of Fig. 2 and also added this sentence to the main text: "The signal integrated in the DI loop is coupled either to the DR loop of another dendrite or the neuronal receiving (NR) loop of the neuron cell body."

Reviewer 2 Comment 3: On page 3, lines 14-16, you may want to clarify that the energy cost of few-photon communication is significant "relative to the energy cost of superconducting electronics used for computation" - or something along those lines. I believe this is the intent of the sentence, but it is not immediately clear from the current wording.

Response: Yes, the reviewer makes a good point. The sentence now reads, "Yet when superconducting electronic circuits are employed for computation, even few-photon communication events represent a significant energy expense relative to the extremely low energy per operation of the superconducting circuits."

Reviewer 2 Comment 1: The paragraph on page 4, lines 32-53, primarily describes biological systems. This may be better contextualized by moving it to a summary paragraph at the end of section 2, linking the description of the technology described in this section back to the potential for use to investigate biological neurons and networks of neurons.

Response: I understand the reviewer's intention here. That paragraph does seem a bit out of place. The paragraph was placed in that location to explain the choice of temporal scales that had been identified in the previous paragraph. To perform this function, I think it is best to leave the paragraph where it is, but to clarify the point of the text, we now begin the paragraph with this sentence: "The reason for focusing on these time scales is as follows." We then proceed to discuss gamma and theta frequencies in biological systems as well as the corresponding frequencies of operation in the hardware under consideration.