We thank the reviewers for their careful reading of our work and for their helpful comments.

Relation to Foldiak and Wiskott et al.'s work and advantages & relevance of a normative approach [R1, R3]: Our paper builds on work by Foldiak [8], who proposed local learning rules to capture relevant temporal correlations, but resorted to numerical simulation to compute the output of the network; and work by Wiskott et al., who proposed SFA as a spectral analysis approach for capturing temporal correlations (or equivalently, slowness) and can be readily computed in the offline setting, but for which there was no efficient multi-channel biologically plausible network implementation. Our normative approach allows us to derive an algorithm with local learning rules directly from the spectral data analysis problem, providing a new perspective on the works of both Foldiak and Wiskott et al. and clarifying their relationship. If accepted, we will expand this discussion in the main text and point out the similarities (and differences) between our learning rules and Foldiak's [8]. We will also clarify that the text in sections 2.1 and 2.2 is essentially a review of work by Wiskott et al.

Relation to biology and experimental evidence [R1]: While we do not map our neural network onto a specific circuit in the brain, the general neural architecture of feedforward and lateral connections is ubiquitous. The main experimental support is from Wiskott et al. who have shown that SFA reproduces response properties of complex cells in the visual cortex and hierarchical SFA reproduces properties of place cells in the hippocampus.

In terms of experimental predictions, our work predicts the synaptic weights in the SFA circuit. For example, it predicts that the singular values of $\mathbf{W}\mathbf{C}_{\bar{x}\bar{x}}\mathbf{W}^\top$ are equal to the singular values of $\mathbf{M}^2$, where $\mathbf{W}$ are the weights of the feedforward connections, $\mathbf{M}$ are the weights of the lateral connections and $\mathbf{C}_{\bar{x}\bar{x}}$ is the covariance matrix of the summed time-lagged signal. With the increasing progress in connectomics, it may be possible to estimate $\mathbf{W}$ and $\mathbf{M}$, which could provide evidence for or against a circuit implementing our proposed SFA network.

With regards to the nonlinear expansion, experimental evidence suggests that the cortex performs (sparse) signal expansions; see, e.g., (DeWeese, Wehr and Zador, 2003) and (Olshausen and Field, 2004). There are various proposed mechanisms for performing sparse codings, including (Olshausen and Field, 1997) and (Arora, Ge, Ma and Moitra, 2014). One mechanism for implementing a *quadratic* expansion are so-called "Sigma-Pi units" (Rumelhart, Hinton and McClelland, 1986), which use gating to implement multiplication of inputs and have been invoked in cortical modeling (Mel and Koch, 1990).

Full rank assumption of the expanded signal [R2]: We assume that the expanded signal $\{\mathbf{x}_t\}$ has full rank for mathematical convenience (namely, so we can take an inverse of the covariance matrix). In reality, the expanded signal may not be full rank. One approach is to first perform PCA on the expanded signal to reduce its dimension; however, this would add a layer to the network. Alternatively, one can simply replace the inverse of the covariance matrix $\mathbf{C}_{xx}^{-1}$ in the objective (see Eq. 8) with its Moore-Penrose pseudo inverse $\mathbf{C}_{xx}^{+}$ (provided the rank of the expanded signal is at least k). In this case, the derivation proceeds exactly as laid out in the paper. We will include a detailed justification in the appendix that this substitution in fact achieves the same objective.

Typos and organization [R4]: Thank you for pointing out the typos. We will correct them and carefully read the paper for other typos and grammatical mistakes. In addition, we will move some previously known technical details to the appendix and expand the discussion of our derivation.