

## Background and Significance

The ability of fMRI to reveal brain activity largely relies on the assumption that the blood oxygenation level dependent (BOLD) signals it measures are approximately proportional to some averaging of neuronal activity over space (several mm) and time (several seconds) [17]. With this assumption, an arbitrary BOLD time course can be linearly de-convolved with a single descriptor, the spatiotemporal hemodynamic response function (HRF), to infer the distribution and time course of neuronal activity [50, 58]. This function describes the stereotypic BOLD signal associated with an impulse of activity, and is of fundamental importance because even slight changes in HRF can result in substantial errors in its ability to predict the timing and location of neuronal activity [25, 136]. Despite its importance, this function has largely been derived by a vast array of indirect and non-simultaneous measurements, such as relating the stimulus responses of individual neurons in anesthetized animals to population level responses across entire cortical areas in awake humans. This diversity of animal models, anesthetic states, stimulation methods, and metrics for quantification, has led to a huge variety of HRF models, and consequently, large controversies regarding what brain signals, neuronal or otherwise, BOLD actually reflects. Perhaps more importantly, recent experiments suggest that HRFs may have been severely mis-characterized due to inadequate temporal sampling. In this proposal, we will address these problem by using a consistent experimental paradigm to simultaneously measure and compare neuronal, metabolic, and hemodynamic responses as a function of stimulus strength, behavioral state, and brain network state using fast optical and MR imaging techniques.

The correspondence between BOLD activity and neuronal activity is still controversial because of a number of factors. First is the lack of consistency in experimental approaches, which given demonstrated variations between individuals, tasks, anesthesia and stimulation parameters in both neuronal and BOLD activation, combined with variations in actual metrics used, makes non-simultaneous comparisons problematic. Most notably in this regard, fMRI experiments typical sample the brain at timescales far too coarse to characterize the rapid activity dynamics that underlie normal cognition and behavior and measured by electrophysiological techniques. Second, there has been a lack of direct measurements of a key intermediate between activity and neurovascular response, namely cellular metabolic activity. Finally, although there is clear evidence that large scale spontaneous activity variations affect neuronal responses [112, 92, 19, 5, 76, 8, 37, 115, 27], BOLD measurements [11, 118, 82, 122], and behavior [21, 2, 99, 33, 68], there have been no attempts to integrate the effects of controlled manipulations of large-scale activity into a computational framework.

Our experimental design addresses all of the limitations (**Fig 1**). Our experiments combine recent technical advances including the ability to functional characterize populations of individual neurons in visual cortex with 2-photon  $\text{Ca}^{++}$  indicators, the ability to image with high spatial and temporal resolution metabolic activity in awake behaving primates, and the ability to acquire BOLD signals with unprecedented temporal resolution. The experiments address, through simultaneous multimodal measurements of brain activity that span multiple spatial and temporal scales, the issues of intersubject, intersession, and intertrial variability which have plagued previous investigations of the neuronal basis of BOLD and will elucidate responses from the levels of neurons to voxels. Critically, we will study a key intermediary in the transformation from neuronal activity to BOLD signals, namely metabolism. Finally, we will use the well established computational theory of divisive normalization to model the effects on network state on event-driven responses.

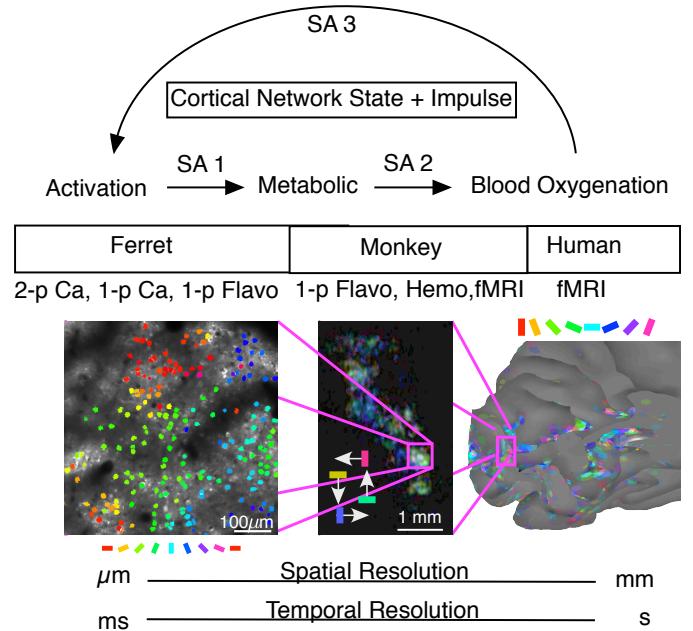


Figure 1: Linking impulse response functions across spatial and temporal scales and variations in behavioral and network state using a consistent experimental paradigm that leverages the functional architecture (orientation selectivity) and response properties (contrast sensitivity and surround suppression) of primary visual cortex. Orientation selective responses obtained by the experimental team are mapped using the three animal preparations employed in this study (ferret, monkey, and human). The first two specific aims are designed to reveal the spatiotemporal dynamics describing key steps in the pathway from orientation-specific activation at the level of single neurons (microns & ms) to the level of blood oxygenation (mm & s). The third specific aim incorporates these results, as well as results relating resting state dynamics to visual responses in humans, to constrain an inverse model going from BOLD to neurons.

**Importance of network and behavioral state** The central challenge that many efforts to compute HRF face is the integration of data sets acquired under different experimental conditions which are known to affect neuronal, metabolic, and hemodynamic responses. One such condition is overall behavior state or level of anesthesia. Although anesthesia results in decreased BOLD activation in animals [39] and humans [3], these results are hard to interpret because anesthesia's effects are not simply a uniform suppression of response. It can produce highly variable changes in different neurons [1] and alter the balance between excitation and inhibition [132]. A further complication is the possibility that anesthesia is modulating the coupling between neuronal activity and BOLD signal by effecting astrocyte function [128] or vasodilation [46]. In addition to behavioral state, HRFs depend on many other poorly understood factors including the individual subject [114], brain region [108, 22, 97], and even voxel that is analyzed [136].

One such important factor are slow (<10 Hz) and distributed (cm) patterns of activity, which we will term network state. As with anesthesia, neurophysiological experiments have demonstrated that network state affects every level of neuronal information processing, from the summation of synaptic potentials in individual neurons [119, 61, 37, 64], to the correlation patterns observed among firing neurons [100] and local field potentials [75] which reflect synaptic potentials, membrane fluctuations, and action potentials [24, 77, 23]. Similarly we know that both BOLD responses [11, 118, 82, 122] and behavior [21, 2, 99, 33, 68] are substantially modulated by network state. Network state might also affect the relationship between BOLD signals and neuronal activity [123].

As with HRF studies, variations between experimental preparations, such as between anesthetized and awake preparations, have made interpreting the effects of network state on event-related activity challenging. For example, although anesthesia has been assumed to simply uniformly suppress brain activity [49, 127], recent studies have revealed that it may have more complicated effects and change the dynamics of large-scale neuronal networks [32, 84, 93]. Also common to HRF studies, different studies have used a variety of definitions for network state, including EEG rhythms [99, 60] and fMRI default mode dynamics [88].

Studies of network state also suffer from two additional limitations that most event-related designs do not: a reliance on spontaneous activity fluctuations without any way to experimentally control the magnitude and nature of those fluctuations, and the lack of a computational framework for explaining how broadly distributed patterns of activity impact local responsiveness. In this proposal we will address both of the issues by making use of the same contrast dependent responses in V1 [104, 52] that led to the establishment of theory of divisive normalization, which quantitatively describes how network state affect responses at a cellular level.

Normalization models have successfully reconciled apparent discrepancies between single unit and BOLD measurements [16, 149] regarding the magnitude of attentional modulation observed in visual cortex [62] and have been validated with direct, although relatively local, activation via optogenetics [101]. Several members of our team have applied divisive normalization models to both electrophysiological [55, 53] and fMRI [73] data. In this proposal, we will employ the original formulation of this theory, in which a divisive signal based on the average level of contrast across all orientations and visual space suppresses the activity to a neuron's preferred stimulus [66]. We will manipulate the network state by varying the contrast of full-field dynamic noise and apply normalization models to predict these effects of these manipulations on activity, metabolism, and hemodynamics. This design carries an additional advantage in that, unlike the very unnatural environment imposed in almost every visual fMRI experiment in which small stimuli are placed on a uniform background, it will create a network state across V1 with completely natural spatiotemporal statistics.

**Importance of metabolism** A pure electrophysiological approach also misses a key intermediate in the transformation of neuronal activity into BOLD responses, namely, metabolism. Neuronal activity is energetically expensive and is sustained by oxidative metabolism in mitochondria. The BOLD response arises because neuronal activation produces an increase in the cerebral metabolic rate of oxygen consumption ( $CMR_{O_2}$ ) [48], yet most in vivo assessments of activation-evoked changes in  $CMR_{O_2}$  rely on tissue or blood oxygenation measurements and, therefore, are indirect. The situation is further complicated by the fact that activity increases both  $CMR_{O_2}$  and cerebral blood flow (CBF), which oppositely influence changes in tissue and blood oxygenation.

These problems can be addressed by the more direct measurements of oxidative metabolism possible with flavoprotein autofluorescence imaging (FAI) of the autofluorescence signals associated with the reduction of a chromophore (FAD/FADH<sub>2</sub>) localized to mitochondrial membranes [28, 41, 110, 78]. As might be expected, simultaneous measurement of blood signals and FAI have revealed that FAI responses are faster [123], more spatially localized [123, 137, 71], and more temporally correlated with neuronal activity [131] than blood volume

or oxygenation signals linked to passing vessels.

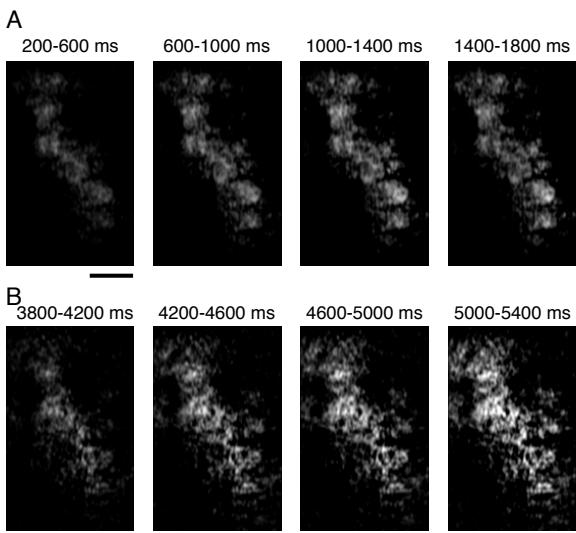


Figure 2: Rapid metabolic increases and slow hemodynamic changes after stimulus onset in awake primate V1. Pixels with a significant ( $p < 0.05$ ) rapid increase in fluorescence (A) show clear orientation column regularity unlike from the pixels with a significant late decrease (B). The late blood-related decrease in response (here inverted to allow for comparison with the positive metabolic response) has both vascular artifacts and a poor spatial correspondence with fast metabolic signals. The limited spatial extent of activation is due to a small stimulus ( $2^{\circ}$  deg). Scale bar=1 mm

Preliminary data attests to the importance of simultaneous measurements of metabolic and hemodynamic effects. In these experiments, monkeys performed a challenging visual detection task during V1 FAI. In accordance with previous studies, we observed a biphasic response to visual stimuli in which a rapid increase in fluorescence was observed within hundreds of milliseconds followed by a extended decrease, which last for on the order of 10 seconds, due to increases in cortical blood volume [131]. Remarkably, while the initial metabolic response is spatially precise, as evidenced by the patchy activation of orientation columns, the subsequent decrease is neither spatially precise nor well aligned with the initial response (Fig 2). These data suggest that the transition from metabolic activity to hemodynamic modulation may be a critical determinant of the precision of BOLD effects.

**Importance of fast measurements** While the vast majority of studies sample BOLD signals with at a repetition time (TR) of 3 to 4 seconds and BOLD latencies can be anywhere between 4 and 14 seconds depending on the brain region and task [69, 81], the BOLD response in a given region, task, and subject has been shown to be impressively time-invariant and temporally precise [29, 94, 30]. In fact, it has been demonstrated that with appropriate paradigms, temporal information over a scale of milliseconds can be resolved with fMRI [103, 87, 30].

Recent studies employing stimuli with fast dynamics have revealed that BOLD responses can follow such dynamics with far greater fidelity than would be expected from the canonical HRF [83] and suggest that traditional block based designs with long TRs are inadequate for fully characterizing BOLD responses [111, 14, 106]. Many recent fMRI studies [30, 44, 126, 130] have demonstrated the utility of faster (sub-second) sampling rates, with respect to both statistical power and detectability of faster components present in the BOLD response [136]. For example, we have found that [30], using sub-second sampling, stimulus shifts on the order of 100 ms have been reliably detected. Recent modeling efforts have shown that it is theoretically feasible with such measurements to deconvolve neuronal activity time-courses from fMRI data [65]. This is a critical need as many important cognitive processes and everyday abilities such as face detection [105] occur over the course of milliseconds.

Ultimately, the goal of this proposal is to understand what neuronal and metabolic signals are being measured by non-invasive human fMRI experiments. However, by employing fast sampling and measuring the physiological responses associated with brief transient inputs, the significance of our proposal extends beyond BOLD interpretation in that it may substantially alter the standard low-TR and block designs currently employed in fMRI studies. Given that intrinsic fluctuations can modulate the amplitude of task-related responses across trials [47, 91] understanding this relationship will also help unite the currently disparate ways of analyzing fMRI data: resting-state, which makes use of slow fluctuations across large populations to infer large scale connectivity [15], and event-driven, which makes use of such event-triggered changes to infer local activation.

Event-driven fMRI applications are currently limited by physiological "noise", which includes all signals of a physiological origin that are not measured. We hypothesize that broadly distributed changes in network state, such as we will measure and manipulate in this proposal, are a major contributor to this "noise" [12]. Such changes, while potentially neuronal in nature, may be unrelated to the evoked response. This is consistent with recent experiments in humans in which ECoG variability is the dominant source of BOLD variability [142] and the notion that network state fluctuations are a substantial contributor to event-related BOLD variability [88]. A full accounting for the nature of the relationship between network state and BOLD responses therefore has the potential of contributing to efforts to reduce or regress out this source of variability, reduce event-related variability, and ultimately increase fMRI sensitivity.

# Innovation

**Parallel cellular, metabolic and fMRI measurements using a common paradigm:** Our proposal uses a consistent paradigm to quantify impulse response functions with high temporal resolution using simultaneous measurements across multiple modalities. This paradigm leverages the functional homology (contrast sensitivity, orientation selectivity, and surround suppression) of visual cortex in ferrets, monkeys, and humans (**Fig 3**).

	Cellular Activity	Meso Activity	Cellular Activity	Metabolism	Oxygenated Hb	Hb Volume	Columnar BOLD	Areal BOLD
Cellular Activity	2-P GCaMP6f	1-P GCaMP6f	2-P JRGECO	1-P 420-480 auto	1-P 530 nm reflect	1-P 670 nm reflect	fMRI Single Voxel	fMRI V1
Meso Activity	1-P GCaMP6f	Exp 1 & 3						
Cellular Activity			Exp 2 & 3					
Metabolism				Exp 4				
Oxygenated Hb				Exp 4				
Hb Volume					Exp 5	Exp 5	Exp 6	Exp 6
Columnar BOLD	MB fMRI				Exp 5	Exp 5	Exp 6	Exp 6
Areal BOLD	MB fMRI							
Whole Brain BOLD	MB fMRI							
	Ferret	Monkey	Monkey+Human					

Figure 3: Simultaneous multi-modal experiments, which combine two or more simultaneous measurements, to link activity, metabolism, and hemodynamics across spatial and temporal scales in columnar cortex.

imaging which allows for measuring, with cellular precision, how populations of neurons respond to an impulse and how those responses are altered by network and behavioral state. The second is the development of metabolic imaging in awake behaving primates based on flavoprotein autofluorescence (FAI) which allows for the direct measurement of event and network state related metabolic responses across a neuronal population. The third is the use of multi-band/simultaneous-multi-slice (MB/SMS) pulse sequences to achieve fast whole-brain functional imaging with sufficient spatial resolution [44, 98, 145] to resolve orientation selectivity at the single voxel level [135]. Our proposal combines expertise in all of these approaches including animal functional imaging (Smith,Ghose), human functional imaging (Yacoub, Kay), high-field MRI engineering (Ugurbil, Adriany, Yacoub), and computational modeling of divisive normalization (Ghose, Kay, Hayden). **Focusing on temporal rather than spatial resolution:** We have found significant orientation tuning at a single voxel level over the scale of mm [135] (**Fig 1**), which is an order of magnitude larger than would be expected on the basis of the size of neuronally defined orientation columns. While numerous fMRI studies aim to achieve sub-mm resolutions to gain further insight into neuronal activities, such studies are plagued by the spatial point spread function of the hemodynamic response suggested by these results. On the other hand, given our preliminary data, there are largely unexplored benefits to pushing temporal resolution, including the possibility of non-invasively measuring the temporal dynamics of neural activity underlying human cognition.

## Approach

### General Methods

**Visual Stimuli and Task** Identical visual stimulation will be used for all experiments. The stimulus consists of a background designed to maintain a constant mean level of activity throughout most of V1, and a flashed probe designed to selective activate subpopulations within V1. The background is a dynamic large-field (>20-30 deg) noise whose spatiotemporal characteristics match natural movies (approximately 1/f in both space and time) [40, 138, 121] at varying levels of RMS contrast (0, 8, 16 %). The probes are brief (200 ms) static flashes of full-field single orientation (0, 45, 90, 135 deg) noise with broad spatial frequency content and varying contrast (4, 8, 16 ,32 %) presented randomly with an inter-stimulus interval between 3 and 30 s with a mean of 10 s (**Fig 4**). The orientation and contrast of these flashes will be randomly interleaved. Background contrast will be changed every 5 minutes in a random order, and each background contrast will be presented at least twice (10 min x 3 contrasts = 30 min) allowing for  $\approx$  60 flashes per background contrast level.

Because all of our experiments will be interleaving two types of observations, we will be able to make quantitative inferences that bridge spatial and temporal scales (**1**). **Cutting-edge technologies:** Our proposal leverages recent technical developments and interleaves these developments to simultaneously characterize orientation selective responses and network state across multiple spatial and temporal domains in primary visual cortex. The first is the development of 2-photon imaging

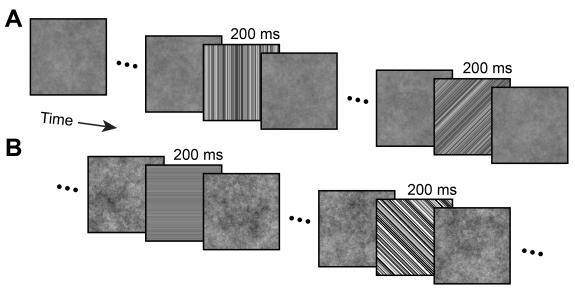


Figure 4: Example trials in which static patterns of a single orientation and contrast are randomly flashed on a background of dynamic noise.

Background contrast will be changed every 5 minutes in a random order, and each background contrast will be presented at least twice (10 min x 3 contrasts = 30 min) allowing for  $\approx$  60 flashes per background contrast level.

A previous limitation in fully characterizing response functions is the long time (on the order of 10s of s) for metabolic and hemodynamic responses to return to baseline [123, 131]. This is problematic for a typical experimental trial-based design for awake behaving monkeys, in which data acquisition and stimulation is turned on and off every few seconds. To avoid this limitation, we will acquire imaging data continuously during the entire 5-periods of dynamic noise. During this period, a fixation point is always present, eye position is monitored, and periods of time for which the subject is not maintaining fixation are discarded in post-hoc analysis. Given the full-field nature of our stimulus, we do not anticipate this to impose a significant limitation since the visual stimulus should always be activating most of V1 for modest eye movements. For monkeys, to reinforce fixation, reward will be randomly administered during fixation. In both human and monkey studies, to ensure vigilance, subjects will be required to hit a button within 750 ms of a single-orientation flash.

**MR Hardware** MRI scans will be conducted on a 7 Tesla Siemens console equipped with the latest generation (SC72) high performance whole body gradient system. Our 7T system has recently been upgraded to 64 receiver channels. Because a high number of receiver channels permits higher accelerations [140] due to significantly lower g-factors compared with 32 channel coils [7]. We will build a 64 channel array coil that has many more coils along the slice direction. This in turn will allow for higher slice accelerations and support faster whole brain TRs for human studies. In the monkey experiment (Exp 6), we will use a specialized RF coil designed to accommodate the optical imaging chamber and the lens/illuminator above the chamber. The close proximity of the RF coil to the receive array will preserve the SNR benefits of tight fitting receive arrays while achieving significantly more efficient RF excitation. For this, a small loop coil will be tightly wrapped around the chamber and positioned as close to the skull as possible. We plan to furthermore increase SNR and accelerated SNR by arranging additional small 2cm loop coils around the small chamber surface coil.

**Univariate and multivariate analyses** Given heterogeneities in neuronal activity, metabolism, and vascularization the assumption that HRFs can be described as simple spatial filters is clearly limited, [79, 4, 117], but addressing this problem requires not only simultaneous multi-modal experiments such as we are proposing, but also methods for analyzing high-dimensional multi-modal data. Although some our data is amenable to classical correlation analyses, such as event-triggered responses, in many cases we wish to analyze the statistical relationships between multimodal data of high dimensionality. Examples of this include responses across a sampled neuronal population vs the spatial pattern of mesoscopic (Exp 1) and metabolic activity (Exp 2). In these cases we will employ a Kernel Canonical Correlation Analysis (kCCA) that computes a multivariate temporal filter to links one data modality to another. In this analysis, a canonical correlogram reflects the coupling dynamics between the two sources and a temporal filter reveals which features in the data give rise to these couplings and when they do so [13]. To test the results of these analyses, we will cross-validate across different background contrast strengths. Of particular interest with regard to the resting state literature, we will test the ability of kernels derived from high background noise contrasts (in which global activity is dominated by the noise stimulus [45]) to predict the effects of spontaneous activity (namely 0% contrast) on impulse responses.

**Consideration of Sex and Other Relevant Biological Variables** Although these experiments quantify responses in a primary sensory system for which there is no evidence of sexual differences, we will record the sex of experimental subjects in all of our experiments as a possible biological variable. Because we use relatively few monkeys, as is the norm for NHP research, we are not able to address issues of sexual differences in this data set. Human experiments (Exp 5 & 7) will include approximately equal numbers of male and female adult volunteers with the only selection criteria being normal vision and a willingness to participate in multiple scanning sessions.

### **Specific Aim 1: Quantify cellular, mesoscopic, and metabolic impulse responses as a function of state**

**Rationale** In order to understand how cortical state affects event related BOLD responses, we must first examine how it affects event related responses at the neuronal and metabolic level. Traditional methods have used extracellular methods to record the event related responses of single neurons in isolation. Such studies ignore not only the context of simultaneous activity patterns among other neurons but also miss signals such as synaptic potentials and membrane fluctuations which may be a significant contributor to hemodynamic responses [63]. Moreover, most single-unit electrophysiological experiments are biased towards measuring output of large pyramidal cells one at a time. To avoid these limitations, we will use a combination of cellular and wide-field imaging

in ferret visual cortex. To measure impulse response functions, we will deliver brief pulses of full field single orientation stimulation (**Fig 4**). We will characterize the spatiotemporal pattern evoked by such pulses and how that pattern depends on both input strength (pulse contrast) and cortical state (background noise contrast).

The ferret visual cortex is well suited to addressing these questions for several reasons: 1) ferrets have a highly columnar visual cortex, similar to organization to other carnivores such as the cat, and to both non-human primates and humans; 2) ferret visual cortex is well suited to viral expression of calcium indicators and both multiphoton and wide-field imaging, approaches which have proved challenging in non-human primates; and 3) ferrets are amenable to head fixation, allowing us to manipulate brain state in a variety of ways.

### Experiment 1: Cellular and mesoscopic impulse response functions as a function of network state

Given that metabolic signals sum responses from local pools of individual neurons, it is critical to understand whether changes in the magnitude of these pooled signals result from changes in the responses of the neurons within the pool or rather reflect a changing makeup of the responsive population. Therefore, in order to determine how cortical state impacts the coupling of neurons into columnar populations, we will examine the responses in identified populations of neurons across a variety of brain states.

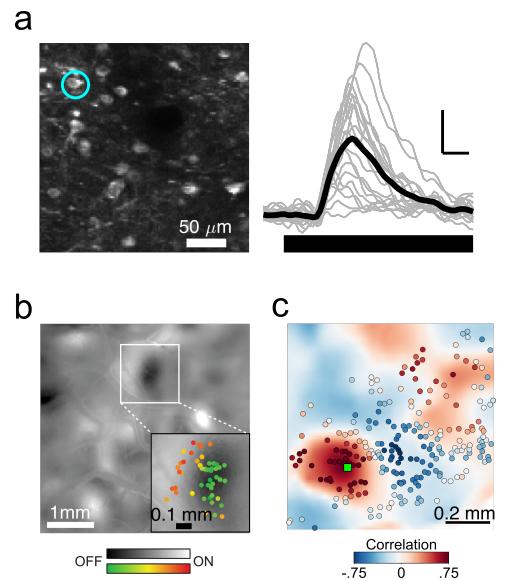
Neurons will be labeled with viral injections of AAV1-hSyn.GCaMP6f (UPenn Vector Core) as described previously. After allowing 2-3 weeks for viral expression, a cranial window will be implanted over primary visual cortex [125, 124]. Responses will be evoked through the brief presentation of full-field grating stimuli, optimized to drive responses in a large proportion of ferret visual cortical neurons. Preliminary data demonstrates that brief visual stimuli evoke transient responses in cortical neurons, and that the rapid onset of the response is well captured by the calcium response (**Fig 5a**). To study neuronal response functions, we will measure how the responses to flashed single orientation probes vary as a function of contrast and background noise contrast (e.g. **Fig 7**). Traditional analyses, as have been applied previously at the level of individual neurons, will be used to characterize whether these contrast dependencies are consistent with normalization models [26, 101, 55].

To quantify the coupling of individual neurons across the sampled population, we will calculate the coherence of the population response to orientation probes on a trial-by-trial basis. Coherence will also be calculated for background stimulation epochs. Of particular interest is the zero contrast background, for this coherency reflects purely spontaneous activity. From these analyses, we can determine the degree to which synchronous populations defined by spontaneous reflections match those populations seen with evoked activity.

To better understand how ongoing activity within the cortex impacts the population coupling of neurons, we will combine wide-field imaging of population activity with simultaneous imaging of individual neuronal responses. Our preliminary data shows that trial-averaged wide-field calcium responses accurately reflect the tuning properties of neurons at that cortical location recorded during a separate imaging period *both for visually-evoked activity* (**Fig 5b**) and *also during ongoing spontaneous activity* (**Fig 5c**). This approach will thereby allow us to capture the ongoing activity patterns of large populations of neurons spanning millimeters of cortex and relate that to the evoked responses in individual neurons.

To achieve this, we will rapidly (>10 Hz) switch between imaging modalities using a custom-built multiphoton and wide-field microscope equipped with a low-power, high NA objective. A fast switching LED will be used to provide excitation for wide-field imaging, and fast shutters will shield the photomultiplier tubes (used for multiphoton imaging) during wide-field acquisition.

By manipulating background noise contrast (**Fig 4**), we will be able to manipulate cortical state and identify how the coupling of individual neurons into a population response is modulated by the pattern of ongoing activity in the cortex. Because of the large number of neurons simultaneously sampled we will apply dimension reduction algorithms to characterize event-related activation dynamics [150]. We will use factor analysis (FA) to separate spiking variability into a



**Figure 5: Visually-evoked responses and ongoing cortical activity.** **a.** Responses to a luminance transition (stimulus onset indicated by black bar) of a neuron indicated to left. Black trace is average over 20 trials, shown in gray. Scale bars: 250 ms, 0.5  $\Delta F/F$ . **b.** Trial-averaged cellular tuning properties are well-matched to those of trial-averaged wide-field population responses imaged separately in time. Main panel shows preference for ON or OFF luminance transitions with wide-field imaging. Inset shows overlaid cellular ON / OFF preference. **c.** Correlations in spontaneous activity of a neuronal population relative to a single cell (green), overlaid on population correlations from wide-field imaging.

component that is shared among neurons and one that is independent across neurons and specifically test the correlations between the shared variability component and mesoscopic measurements.

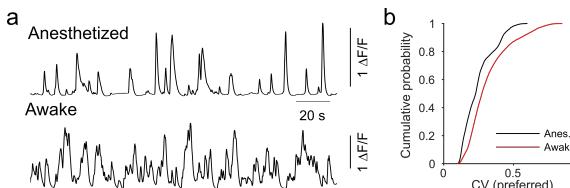
In addition to the measures of population coupling outlined above, for each neuron we will create maps triggered on high cellular activity [144] to directly visualize the spatial motifs at a mesoscopic scale that are highly coupled to neuronal firing. To identify whether certain states are more likely than others to influence responses to our flashed orientation inputs, we will use machine learning algorithms to identify and classify ongoing cortical activity states. We will then determine how visually evoked responses and population coupling are modulated by particular mesoscopic states. **Hypothesis** We hypothesize that cellular level responses will be well described by a normalization model which incorporates both stimulus input strength and mesoscopic measurements of neuronal activation.

## Experiment 2: Cellular and metabolic impulse response functions as a function of network state

Increased metabolism at a cellular level is a critical intermediary between activity and the hemodynamic changes measured by BOLD, but there have been relatively few experiments which have directly measured this via flavoprotein autofluorescence imaging (FAI) and none which have directly mapped the relationship between activity and metabolism with cellular resolution. To address this shortcoming, we will interleave 1-p measurements of FAI with 2-p  $\text{Ca}^{++}$  imaging. Because the GCaMP6f vector would be saturated by 420-480 nm light used for FAI [36], we will employ a red-shifted  $\text{Ca}^{++}$  indicator such as jRGECO1a. As with the 1-p metabolic measurements, we aim to understand how metabolism depends on input strength (orientation contrast) and background activity (background contrast). Similar to the aforementioned analyses, we will construct cellular activity triggered maps to understand spatial motifs of metabolic activity, and cross-validating our models of the relationship between cellular activity and metabolism by predicting cellular level activity on the basis of metabolism and metabolism on the basis of cellular level activity. **Hypothesis** We hypothesize a tight relationship between cellular metabolism and activity such that the conclusions of Experiment 1 regarding the normalization of activity can be extended to the domain of metabolism.

## Experiment 3: Cellular, mesoscopic, and metabolic impulse response functions as a function of behavior state

In the experiments described above, we will use visual stimuli to modulate the level and structure of ongoing cortical activity in an anesthetized animal. These experiments are essential to quantify the coupling of neurons into populations and the impact on metabolic responses. However, preliminary data demonstrates that the pattern and structure of ongoing cortical activity can differ dramatically between behavioral states, with much higher frequency events appearing in the awake cortex relative to anesthesia (**Fig 6a**).



**Figure 6: Visually-evoked responses and ongoing cortical activity.** a. Ongoing population activity measured with wide-field imaging in the anesthetized (top) and awake (bottom) states. b. Trial-to-trial variability of evoked responses is significantly higher in the awake state.

Thus it is likely that ongoing activity in the awake cortex may play an even greater role in modulating the response evoked by a visual stimulus. Indeed, preliminary data suggests that, at the population level, trial-to-trial variability is heightened in the awake state, suggesting a greater effect of ongoing activity (**Fig 6b**).

To test this, we will manipulate behavioral state by varying the level of anesthesia, and by comparing awake animals to those under anesthesia while presenting visual stimuli as described above and assaying response variability at both the single neuron, population, and metabolic level. In order to control for response variability induced by eye movement, we will use large, full-field stimuli

which will mitigate the effects of variations in the positioning of receptive fields. We will also employ eye-tracking, and will exclude periods of large eye movements from further analysis. **Hypothesis** We hypothesize that the fundamental relationships observed in Experiments 1 and 2 will be validated, and that the average level and variability of activity and metabolism will be higher in the awake animal.

**Potential Pitfalls** Calcium imaging indirectly reports neuronal activity, with spiking activity being convolved with much slower kinetics for both calcium and reporter fluorescence. We will overcome this by utilizing brief visual stimuli, as the our preliminary data indicates that the initial rise of fluorescence is rapid and readily detected (**Fig 5a**). In addition, we will employ standard deconvolution approaches to estimate underlying spike trains from calcium signals [133]. An additional potential pitfall is the choice of ferrets as an animal model, as the coupling

of neuronal activity to population and metabolic activity in ferrets may differ from that in primates. However, the ferret visual system has been extensively studied and it shows a clear columnar architecture, similar to that in primates. In addition, ferrets permit high quality multi-photon imaging (e.g. [125, 124]) which is difficult in primates, and facilitate a greater experimental throughput than experiments in primates. Furthermore, by employing the same stimulus set across species, while also overlapping methodologies (metabolic imaging) we will be able to verify that the activity patterns are consistent across species. Lastly, it may be the case that mesoscopic imaging captures fundamentally different signals than those at the cellular level which have been shown to primarily reflect spiking activity. Although our preliminary data indicates that the tuning preferences of neurons are well-matched to the properties obtained with wide-field mesoscopic imaging of the same cortical location (**Fig 5b,c**), strongly suggesting that wide-field signals also reflect the spiking properties of neuronal populations, any discrepancies could fundamentally limit the spatial precision of BOLD measurements, and therefore be important to quantify.

## Specific Aim 2: Quantify metabolic and hemodynamic impulse responses as a function of state

**Rationale** BOLD responses are linked to neuronal activity via activity-related increases in metabolic load, yet the dynamics of this critical intermediary have rarely characterized at high spatial and temporal resolution. To address this limitation we will use our visual full-field stimulation paradigm to obtain spatiotemporal maps of the metabolic impulse response function in awake behaving non-human primates using fully MR compatible optical imaging methods.

As with the ferret preparation, by presenting brief flashes of single orientation stimuli, the functions will be measured by delivering both spatially sparse (because of orientation columns), and temporally sparse input. To directly link such responses to the oxygenated blood signal measured by BOLD, we will simultaneously map the dynamics of hemodynamic impulse response functions as well. As with all other aims, by manipulating network state across V1 through the use of full-field dynamic noise of varying contrasts, we measure how these functions vary with network state, and, importantly, how the relationship between metabolism and hemodynamics depends on network state. Critically, these experiments will be conducted in an animal whose visual cortex is highly homologous with human visual cortex, including features such as cytochrome oxidase periodicity [143, 120] which may be relevant when examining the spatial pattern of evoked metabolic transients.

**Surgical procedures for non-human primates** Before any training, high-resolution (0.5-mm isotropic) anatomical brain images will be acquired by securing the anesthetized animal in an MR-compatible stereotaxic frame inside a 7 T MR scanner. High-resolution images of the skull will be acquired using a CT scanner and be used to fabricate a customized polyether ether ketone (PEEK) headpost which is secured to the animal's skull in a sterile surgery. Once the animals are trained to perform the tasks, a second surgery is performed in which a V1 craniotomy/durotomy is performed. V1 is identified anatomically by visualization of the operculum of V1 (ensuring that the exposed region of V1 is not near the lunate sulcus and V2/V4) and by cranial landmarks. To maintain patency of the craniotomy and durotomy, a biocompatible polymer (polyethyl ether ketone, PEEK) imaging chamber and transparent artificial dura [6] is placed over V1. All materials are MR compatible: the chronic skull implant is custom fabricated in PEEK according to the individual animal's skull contours as obtained by CT scan. At the conclusion of all imaging experiments, animals will be euthanized and V1 tangentially sectioned and stained for the metabolic enzyme cytochrome oxidase (CO) [143, 120]. Histological slides will be aligned to optical images using vascular landmarks and CO blobs will be registered with FAI maps to study the relationship between metabolic capacity, event evoked metabolism, and functional organization [38, 80].

**Task** After recovery from chamber surgery, animals will fixate while viewing our dynamic noise + orientation pulse stimuli. Images will be acquired from a CMOS camera suspended over the animal's exposed V1 and held rigidly to the behavioral apparatus. The camera will acquire a 256x256 pixel image spanning  $\approx 8 \times 8$  mm of the cortical surface at 10 Hz (23.4 $\times$ 23.4  $\mu\text{m}/\text{pixel}$ ). For autofluorescence exposures, the chamber is illuminated with blue light (415-485 nm band-pass Chroma D455/70x filter), and for hemodynamic exposures, red light (570 and 630 nm) is used. Illumination filters are shifted by a shutter synced with acquisition to allow for interleaved acquisition at 5 Hz [123]. Light entering the camera is filtered (520 nm long-pass Chroma E515LPv2 filter) before amplification and 16-bit digitization. Visual stimuli will be delivered using a luminance-calibrated LCD monitor (120 Hz) with an attached photodiode attached to confirm the precise timing of stimulus events. Software designed by the PI (Lablib, OS X) will be used for stimulation, behavioral control, and data acquisition for recordings from visual cortex of behaving rhesus monkeys. Experiments will be conducted while animals are head-fixed.

Animals will be trained in a horizontal position in a MR-compatible primate chair outside of the scanner using operant conditioning with drops of Gatorade or water as rewards for successfully completed trials ( $\sim 0.1$  ml). The animals' fluid intake will be constantly monitored to ensure well-being and adequate motivation for the task. Accommodation to scanner acoustic noise will be achieved using audio recordings of gradient sequences. Eye position will be monitored using an IR camera (Eye-Scan) and body position will be monitored via a Moire Phase Target (Metria Innovation) positioned on a harness secured to the animal's torso. Within the magnet behavioral trials will be TR synchronized [107, 89, 74, 59, 10].

#### Experiment 4: Metabolic and hemodynamic impulse responses as a function of network state

After animals are trained to fixate and respond to the orientation flashes, they will be implanted with a chronic optical imaging chamber and transparent artificial dura and imaged on a daily basis outside of the magnet. In these experiments we will simultaneously examine with high spatial resolution (on the order of tens of microns) evoked metabolism, how that metabolism is altered by network state (as manipulated by background contrast), and the corresponding hemodynamic effects. To accomplish these simultaneous measurements we will use a shutter to switch between different illumination frequencies at 10 Hz (resulting in FAI and hemodynamic imaging at 5 Hz) [123]. This sampling is sufficient to fully characterize the dynamics of these signals (Fig 2 & Fig 7).

Hemodynamic responses to activation are known to incorporate changes in both the oxygenation and total volume of blood delivered. We will examine the spatiotemporal dynamics of both of these changes by, in separate experiments, changing our hemodynamic filter from 570 nm (at HbR and HB02 have equal reflectance) to 630 nm (at which reflectance changes are dominated by the oxygenation of hemoglobin) [51]. We will thus be able to systematically study the spatiotemporal precision by which blood volume and oxygenation change as function of input strength (contrast). On the basis of studies done in anesthetized rats [131], we expect that these measurements will reveal a clear temporal sequence in which cellular metabolism changes occur within 200 ms, followed by blood deoxygenation changes several seconds later (the "initial dip" or early BOLD signal [42, 136], followed by blood volume changes several seconds later (the positive BOLD signal associated with increases in oxygenated blood).

As with the ferret, we will test whether normalization models can explain how local metabolic and hemodynamic responses incorporate local and global inputs. Consistent with the tight correspondence to neuronal activity suggested by previous studies [70, 123, 131], in preliminary experiments conducted in awake behaving monkeys, we have found contrast and size dependencies of FAI signals consistent with electrophysiological recordings of single cells and such normalization models [9, 26]. Specifically we find the existence of strong, but temporally delayed, surround suppression (Fig 7A) and differences between the contrast sensitivity of the surround and that of the center (Fig 7B).

Because the FAI and hemodynamic data sets have identical spatial and temporal sampling, classical correlation methods can be applied to measuring the transform between the two maps. Of particular interest, given our preliminary data of a spatial and temporal mismatch between metabolic and hemodynamic changes (Fig 2), we will investigate whether this mismatch is consistent across all the levels of local and global activity evoked by our manipulations of contrast. **Hypothesis** We hypothesize that coupling of hemodynamic to metabolic changes is a major bottleneck in the spatiotemporal precision of BOLD responses with hemodynamic responses being both delayed and spatially distorted with respect to metabolic changes.

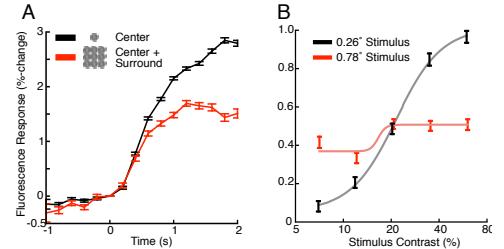


Figure 7: Surround suppression is visible in metabolic imaging experiments in the awake monkey, and consistent with previous electrophysiological results and normalization models. Flavoprotein autofluorescence signals over a  $0.30\text{ mm}^2$  area associated with a small oriented patch rise within 200 ms after stimulus onset (time=0) (A), and the signal is delayed and reduced when the oriented patch is surrounded by similar patches. Around 1 second, these signals display contrast response sensitivities as a function of stimulus size (B) that are also consistent with electrophysiology and normalization models. Error bars=95%SEM

**Potential Pitfalls** Given the electrophysiological [31, 139] and fMRI [148, 96] evidence that brief transients of 200 ms can produce detectable signals that are more robust than would be expected on the basis of traditional block designs, we do not expect any problem resolving metabolic or hemodynamic responses to high contrast flashes. Our preliminary data also suggests that we can obtain responses at lower contrasts as well (Fig 7B). However, in these experiments, as would be expected, the best responses were observed for stimuli that was sized according to the receptive field size, while responses for larger stimuli (such as we proposing to use here (Fig 4) were suppressed. Our stimuli are more transient than those employed in our preliminary experiments, so

it is likely, given the differences in the summation properties of the surround [9], that our responses will not be as significantly impacted. However, if initial experiments indicate otherwise, we will adopt a strategy that we have employed previously to deal with strong surround suppression: the presentation of an array of appropriately sized stimuli with gaps between the stimuli [54, 56].

### Specific Aim 3: Quantify hemodynamic and BOLD impulse responses as a function of state

**Rationale** Accurate characterization of the time-course of the BOLD response is critical to any fMRI experimental paradigm that attempts to estimate temporal dynamics of neural activity from fMRI data. The aforementioned experiments will generate a broad multi-modal data set with which to test models that relate cellular-level measurement of activity to population-level measurements. In this aim we will integrate data generated by the first two aims to inform and validate dynamical models of the relationship of BOLD to neuronal activity.

#### Experiment 5: Optimize temporal fMRI SNR in humans and monkeys

Extraction of such fine-scale temporal information from fMRI measurements requires a method for fast acquisition of images. Traditionally, echo planar imaging (EPI) [90] or spiral acquisitions [57, 95, 102] have been employed in fMRI to obtain fast imaging speeds. However, even using these schemes, readout times for whole-brain volumes can be prohibitively long (multiple seconds). To counter this, one approach is to reduce the acquisition time of a single read train by reducing the imaging field-of-view. Examples of this include outer volume suppression (OVS) [106, 113], smaller coils that are sensitive to smaller volumes, and inner volume excitation [42, 43, 146, 147]. Such limited-volume approaches preclude whole-brain imaging and hinder the effectiveness of motion correction procedures. Other techniques, such as dynamic functional magnetic resonance inverse imaging or MR-encephalography [67, 85, 86, 72], while potentially very fast, are severely compromised in spatial resolution and are associated with variable imaging point-spread function, especially in the center of the brain.

To address these challenges, we have pioneered the use of MB techniques in which multiple slices are simultaneously excited. For human fMRI, 2D MB-EPI has permitted 8-fold (or more) faster volume TRs[44, 98, 145] compared to conventional 2D EPI, without compromising spatial resolution or volume coverage. This is possible because MB slice accelerations do not come with a k-space undersampling SNR penalty that is seen with conventional parallel imaging. Additionally, the application of controlled aliasing to MB-EPI [18, 116] spreads the aliasing energy, resulting in significantly reduced g-factors.

Three criteria will be used to evaluate the different fMRI protocols. First, we will compute L/g factors [129, 145] in order to assess residual aliasing and to minimize g-factor noise amplification. Second, we will compute temporal SNR associated with the different acquisitions and compare results across subjects. Finally, ICA decomposition of the resting-state data will be performed, and common resting-state networks (RSNs) will be identified [35]. The ability to identify RSNs provides a marker of acceptable temporal SNR levels.

We will initially acquire resting-state fMRI (rfMRI) data for 10 subjects, parametrically varying spatial resolution from 1.5 to 2.5 mm, MB factor from 2 to 6, and number of slices from whole-brain coverage to 3/4-brain coverage. In each case, the TR will be set to the minimum achievable value. MB/SMS accelerated fMRI is essential for achieving large coverage of the brain while maintaining relatively fast temporal sampling rates. Preliminary

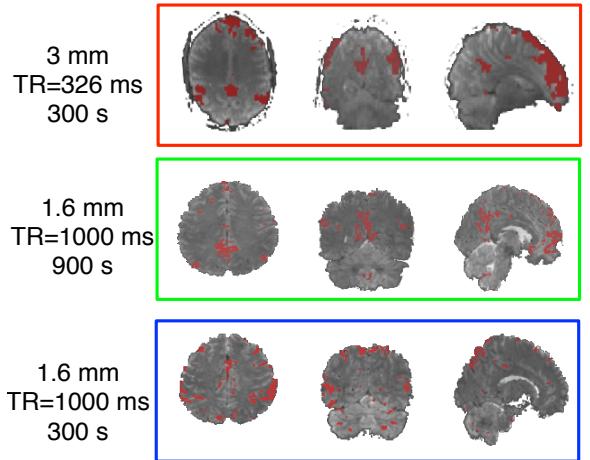


Figure 8: The visual-parietal network is only visible in single subject data acquired over 5 minutes with fast temporal sampling.

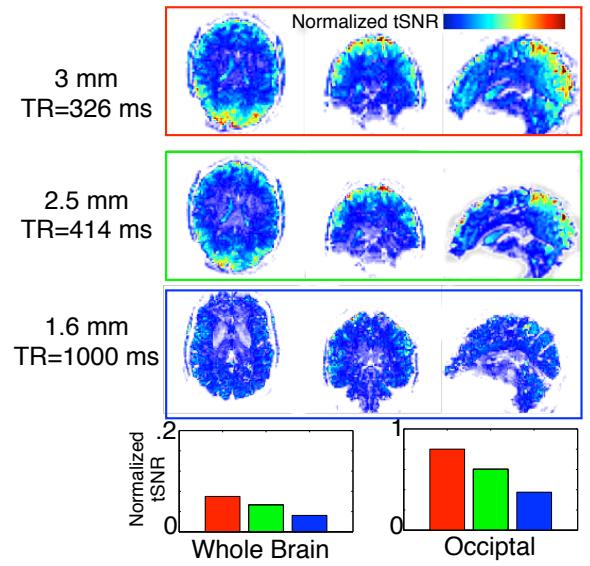


Figure 9: tSNR at 7T with <1 s TRs.

data using MB-fMRI show that low TR acquisitions are particularly advantageous for visualizing default mode networks (**Fig 8**) and characterizing the networks underlying complex behaviors [134]. We will conduct systematic measurements of the impact of MB acceleration, spatial resolution (voxel size), and repetition time (TR) on signal-to-noise ratio (SNR) and temporal SNR to guide MR protocol selection for subsequent studies. The measurements will be conducted using the new 64 channel coil. With the 32 channel coil we have demonstrated whole-brain coverage at 1.5-mm resolution using MB factors of 3-5 (IPAT=2), yielding TRs of 0.6-0.8 s. By reducing spatial resolution to 2.0 or 2.5 mm, we were able to substantially reduce TRs to 0.2-0.4 s (**Figure 9**). The 64-channel coil will increase sensitivity in the cortex as well as the possible acceleration that can be achieved [141].

### Experiment 6: Hemodynamic and BOLD impulse responses as a function of network state

To directly test the assumption that BOLD signals on the scale of mm are a simple spatial averaging of hemodynamic changes occurring on the scale of tens of microns we will optically image hemodynamic changes using the exact stimulus and task paradigms as Experiment 4 in a 7T scanner. We directly examine the controversial issue of the basis of single voxel orientation tuning [79, 117], and the extent to which such tuning reflects columnar organization at the neuronal level (as measured in Exp 4) or vasculature heterogeneity.

After several months of data are acquired for each monkey in Experiment 4 outside of the magnet, we will spend one month accommodating the animal to the magnet environment, and begin multimodal optical/fMRI experiments. A flexible fiber optic light cable will be brought out of the magnet bore and connected to our CMOS camera and illuminator [34]. Using standard retinotopic mapping stimulus, we will first map V1 and verify that our dynamic noise stimuli activates most of V1 (**Fig 10**). We will first optimize the fast TR protocol, as in the human. With the custom built monkey array coil, we will be positioned to achieve 0.2-0.4 s TRs with a 1.5 - 2.0 mm resolution and 0.6-0.8 s TRs with 1.0 - 1.5 mm resolutions, using MB factors of 3-5 and IPAT =2, since the monkey brain and coils are proportionally smaller than the human. We will conduct systematic measurements of the impact of MB acceleration, spatial resolution (voxel size), and repetition time (TR) on signal-to-noise ratio (SNR) and temporal SNR to guide MR protocol selection for subsequent studies. After this we will acquire task data under different pulse and background noise contrasts. We will then concentrate our analysis on the voxels containing the region imaged optically. Of particular interest, given our finding of the existence of single-voxel orientation tuning within human visual cortex, is whether the orientation tuning of a single voxel can be predicted on the basis of vascular heterogeneities [79, 117]. **Hypothesis** We hypothesize that BOLD signals will largely reflect hemodynamic heterogeneities within a voxel (e.g.**Fig 2**) and that normalization rules will be able to describe the effects of contrast on hemodynamic and BOLD responses.

### Experiment 7: BOLD impulse response functions as a function of network state

While it is widely known that a number of factors, such as cognitive state, cortical region and input stimulus strength, have a dramatic impact upon the time-course of the BOLD signal, fMRI studies traditionally ignore these modulations, favoring the utilization of a standard generalized HRF. Here we aim to quantify the extent to which network states and input strength modulate the BOLD time-course and whether these factors impact upon the neuronal-BOLD coupling. As with the previous aims, we will use pulse and background contrast manipulations to measure the extent to which input strength and ongoing activity modulate the HRF and how accurately these HRFS can explain BOLD signal variance compared to canonical HRFs. We will further test whether normalization models can be used to explain single voxel responses as a function of the BOLD signals seen in surrounding voxels[20]. We will also conduct resting state analyses on the zero-contrast background noise to see how activity

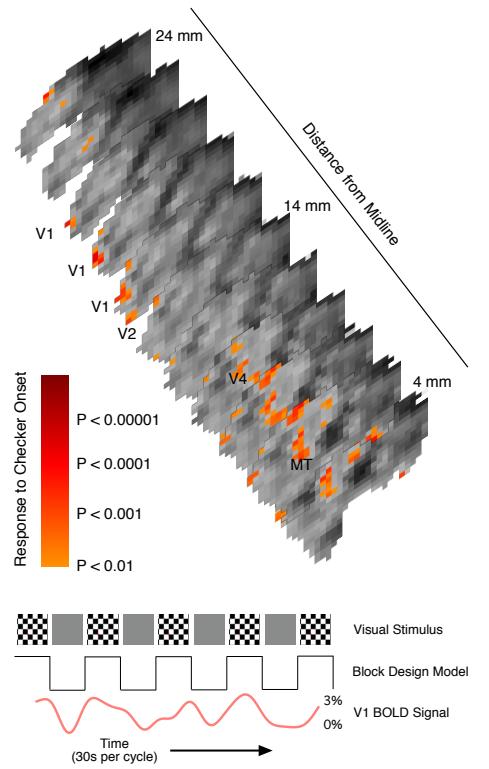


Figure 10: Whole brain fMRI mapping (TR=1s, 2 mm isotropic) of visual activation in the anesthetized non-human primate. Voxels showing visual responses to a 4Hz counter-phasing plaid (one-sided linear regression T-test, 596 DOF,  $p<0.01$ ) were found in early (V1, V2) and intermediate visual areas (V4, MT, other unlabeled clusters) throughout the occipital and parietal cortex. Locations of visual areas was estimated using anatomical landmarks including the operculum, lunate sulcus, and superior temporal sulcus.

in other parts of the brain follows V1 activity and whether these default mode covariations, such as parietal areas, can be used as a proxy for the "average" input that goes into divisive normalization. In all of these cases we will cross-validate on an independent portions of the data set to avoid over-fitting. **Hypothesis** We predict that network state and input strength will have dramatic impact upon the BOLD-time course and that these modulations can be explained using divisive moralization at the single voxel level.

### Experiment 8: Relationship of BOLD dynamics to neuronal activity dynamics

Based on the wealth of data collected across species and modalities, we will develop a simple model that maps single voxel BOLD responses onto population level activity and vice-versa. We will infer the dynamics of task-relevant population level neuronal activity by deconvolving the BOLD data acquired during task performance (Exp 7) with measured impulse response functions (Exp 1-6). We will validate this approach by, on a single voxel level, by trying to replicate the response dynamics observed in the previous aims and how those dynamics change with contrast manipulations. Finally, we will compute a whole brain map that will summarize the fidelity of these cross-modality activity predictions, indicating cortical regions where such prediction fails, thus providing a window onto cortical regions or network states that will require further investigation. **Hypothesis** We hypothesize that, based on the BOLD response of a single voxel and the activity-metabolic-hemodynamics relationships elucidated in the previous aims, we will be able to generate realistic neuronal and metabolic activity impulse responses, and accurately replicate how these responses vary with pulse and background contrast.

**Potential Pitfalls** Our preliminary data show that, with non-optimized coil designs and under anesthesia, we can obtain low latency BOLD signals (**Fig 10**), so we do not anticipate problems in this regard. Exp 7 will depend on observable BOLD responses to pulsed stimulation, and accordingly may require titration of task parameters as described in Aim 2. Given our previous observations of single voxel orientation tuning [135], we anticipate that, at least for some voxels, we will be able to generate realistic response predictions. Should this not prove true, we will use a multi-voxel approach called temporal multivariate analysis that has been shown to retrieve temporal differences between conditions that would otherwise go unnoticed using classic univariate approaches [109]. Should large number of voxels fail to generate realistic activity and metabolic response dynamics with our linear "inverse" filter approach, we concentrate our efforts into developing non-linear cross-modality models.

## Milestones

Technical developments necessary for the human (pink) and monkey (cyan) experiments will be completed in the first 6 months (Exp 5) and computational integration of the human and animal experimental data will be done in the final 12 (Exp 8). Total subject numbers are 48 humans (pink), 6 monkeys (cyan), and 18 ferrets (green). All monkeys will be used 2 at a time in the following sequence: training (2-4 months), optical imaging (2-4 months), optical + fMRI imaging (2-4 months). Previous experience in our lab has demonstrated that our chronic optical imaging chambers remain viable for over a year. At the conclusion of the imaging studies, animals will be euthanized to allow for sectioning and staining for the metabolic enzyme cytochrome oxidase. Half of the ferrets will be used in the initial anesthetized experiments (Exp 1 & 2) and half will be chronically implanted to allow for both awake 3 and anesthetized experiments in the same animal.

		Ferret Monkey Human	Year 1	Year 2	Year 3	Year 4
Specific Aim 1	Exp 1	2-P/1-P GCaMP6f (N=4)				
	Exp 2	2-P jRGECO & 1-P 420 nm (N=5)				
	Exp 3	Exp 1 and Exp 2 in awake (N=9)				
Specific Aim 2	Exp 4	train monkeys (N=6) 1-P 420 & 1-P 530/670 (N=6)				
	Exp 5	MB/SMS Development				
Specific Aim 3	Exp 6	optical imaging and MR coil/environment integration (N=6) 1-P 570/630 & BOLD (N=6)				
	Exp 7	acquire fast fMRI data on humans (N=48)				
	Exp 8	develop inverse models on basis of Exp 1-6				