

SUPPLEMENTAL INFORMATION

Simultaneous multi-plane imaging of neural circuits

Weijian Yang¹, Jae-eun Kang Miller¹, Luis Carrillo-Reid¹, Eftychios Pnevmatikakis^{2,3}, Liam Paninski^{1,3}, Rafael Yuste¹, and Darcy S. Peterka¹

¹Neurotechnology Center, Department of Biological Sciences, Columbia University, New York, NY 10027, USA

²Simons Center for Data Analysis, Simons Foundation, New York, NY 10010, USA

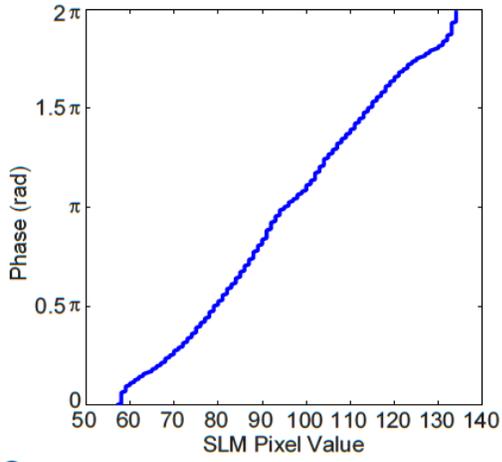
³Department of Statistics, Center for Theoretical Neuroscience, and Grossman Center for the Statistics of Mind, Columbia University, New York, NY 10027, USA

Supplemental Figure S1	SLM defocusing and deflection efficiency (related to Experimental Procedures)
Supplemental Figure S2	Signal extraction using CNMF (related to Fig. 4)
Supplemental Figure S3	Comparison of CNMF and ICA performance on signal extraction from single-plane population calcium imaging (related to Fig. 4)
Supplemental Figure S4	Correlation between signals extracted from CNMF and non-overlapped source regions (related to Fig. 5)
Supplemental Figure S5	Pooled statistics of CNMF performance over 10 different dual plane imaging sessions across 5 mice (related to Fig. 5)
Supplemental Figure S6	Towards fast volumetric imaging: simultaneous axial three plane imaging extending the axial plane separation over 500 μm , and time-multiplexed strategy for SLM-based volumetric imaging (related to Fig. 7)
Supplemental Table S1	Zernike polynomials and coefficients for defocus, first-order spherical aberration and second-order spherical aberration (related to Experimental Procedures)
SLM Beam Steering (related to Experimental Procedures)	Maximum lateral beam steering
	Maximum axial beam steering
	SLM pixel number, pixel size, and fill factor
	Steering dependence on objective
	Chromatic Effects
	Adaptive optics
	Power delivery to each axial layer
	Hologram generation
SLM switching speed	

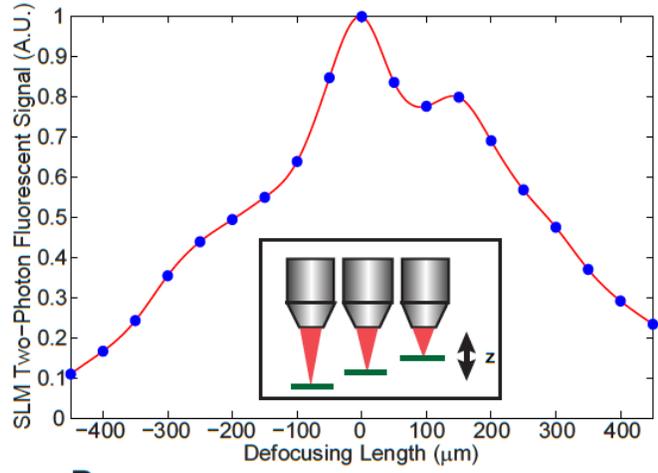
	Beam configurations for multiplane imaging
Source Location Initialization in CNMF Algorithm (related to Experimental Procedures)	
Supplemental References	

Supplemental Figure S1 | SLM Defocusing and Deflection Efficiency

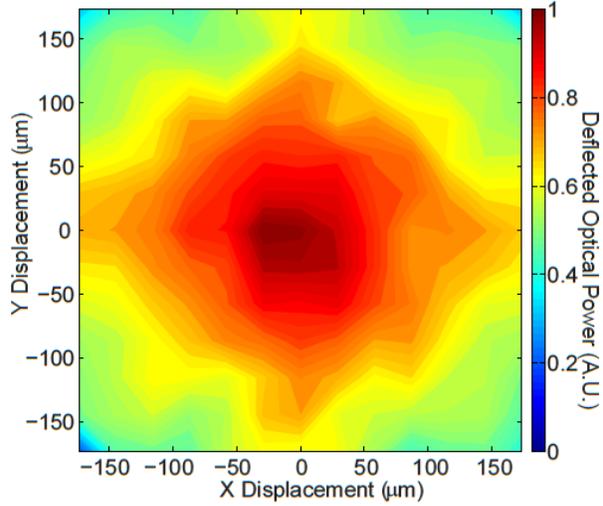
A



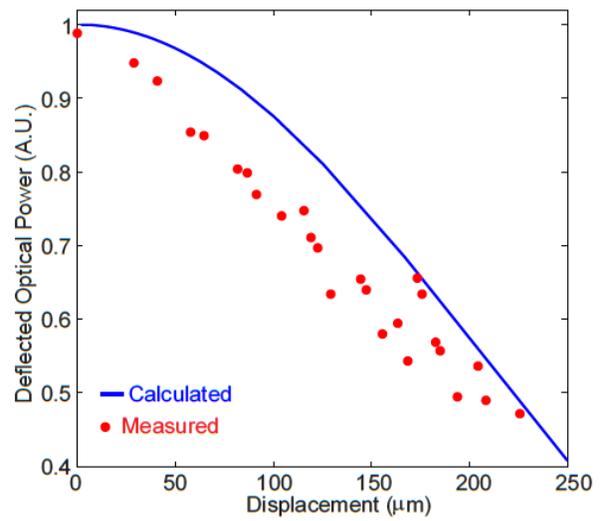
B



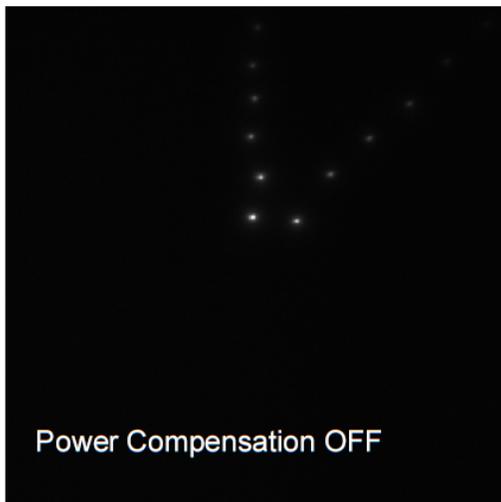
C



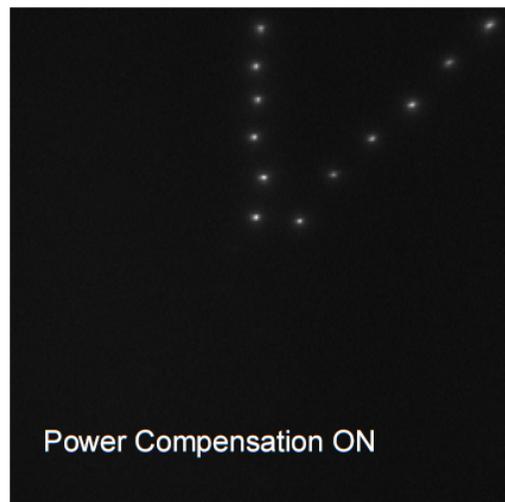
D



E

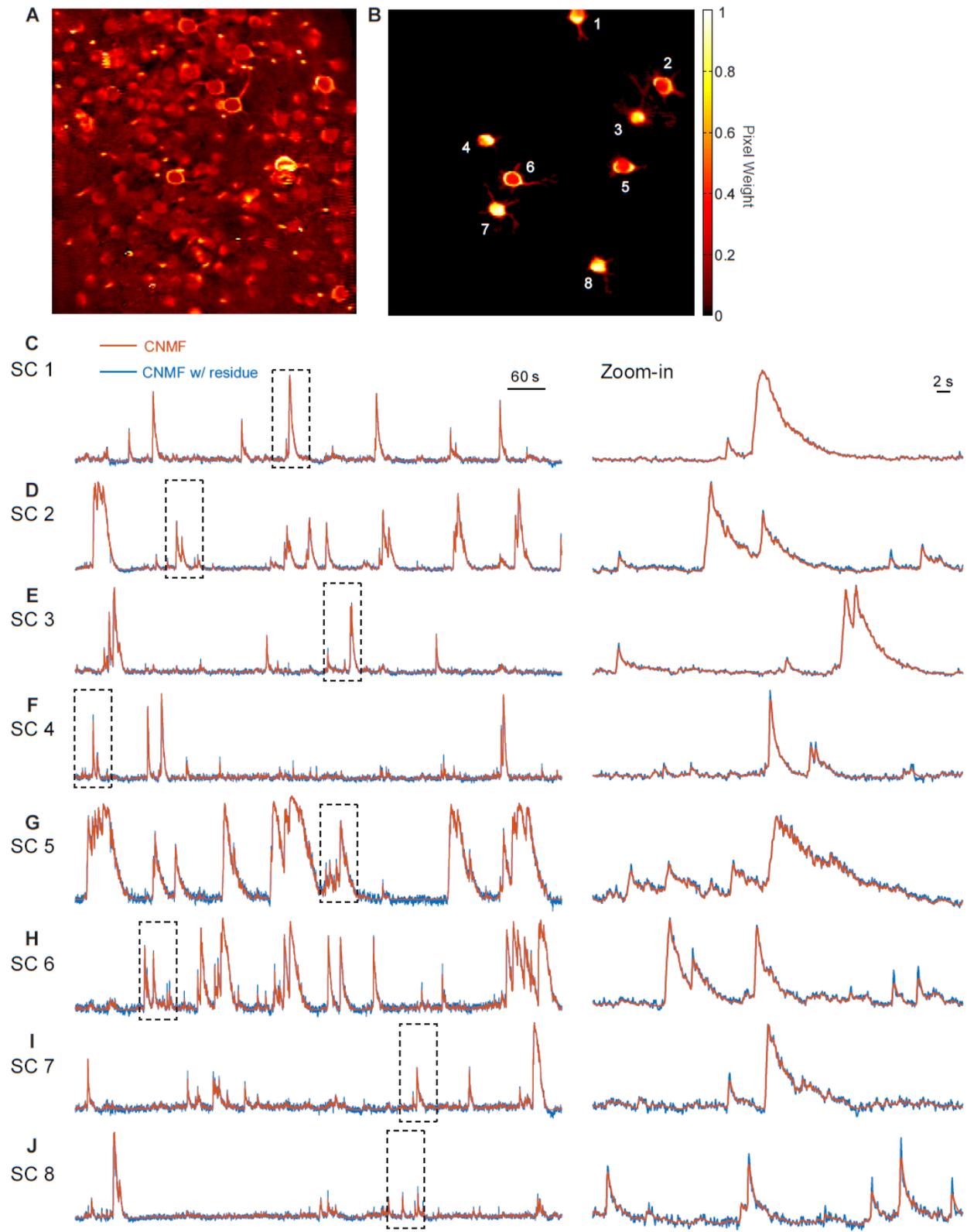


F



- (A) The applied driving pixel value to the SLM (256 steps in total) versus reflected phase of SLM for wavelength of 940 nm.
- (B) SLM two-photon fluorescence efficiency with different defocusing length, measured from the fluorescence emitted from thin Rhodamine 6G sample (50 μm) with two-photon excitation. The experimentally measured data (blue dots) is fitted with red curve, with a full-width-at-half-maximum of $\sim 500 \mu\text{m}$.
- (C) SLM lateral deflection efficiency, measured at the back aperture of the objective. We indexed the FOV with a 10x10 grid, and used the SLM to sequentially target each of the grid points. The deflected optical power is captured by a calibrated photodetector (Coherent FieldMax II with PM10X head) and linear interpolation was used to create the display surface.
- (D) Calculated SLM lateral deflection efficiency (blue curve), with measured data from (C), averaged radially, shown as filled red circles.
- (E) Two-photon fluorescence from a uniform sample, illuminated by twelve simultaneous beamlets generated by the SLM, imaged by an EMCCD (Andor iXon897), without power compensation for SLM lateral deflection efficiency.
- (F) Same as (E), but with compensation to adjust for SLM lateral deflection efficiency (gray level hologram with intensity weighting scaled to compensate the measured fall-off). The intensities of the different spots become nearly uniform after this compensation.

Supplemental Figure S2 | Signal extraction using CNMF

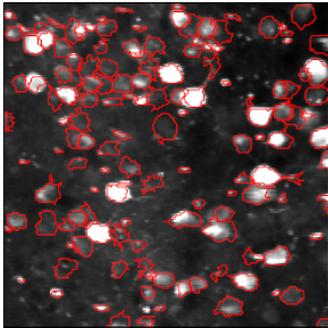


Signal extraction using CNMF, showing the diversity of the signals (various shapes of the spatial component including doughnut shape and perisomatic dendritic processes, as well as various calcium dynamics in the temporal signal) that CNMF can extract. We note that though many transients with high amplitude and long decays are detected, the same neurons also show sharp, lower amplitude, fast time decay transients. CNMF extracts both types of signals well.

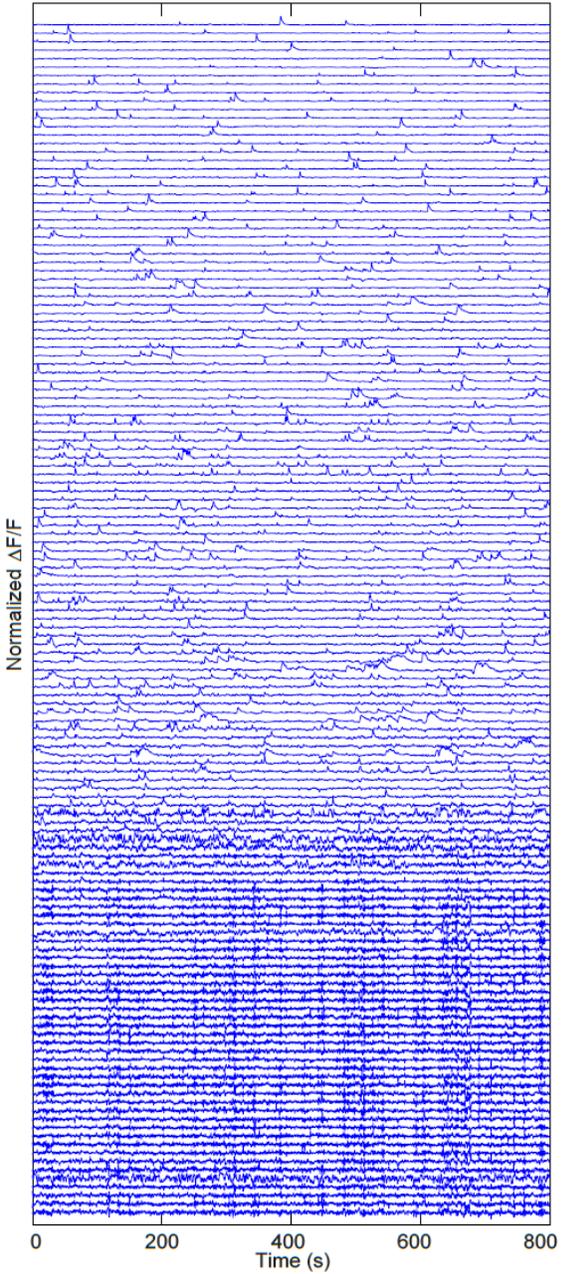
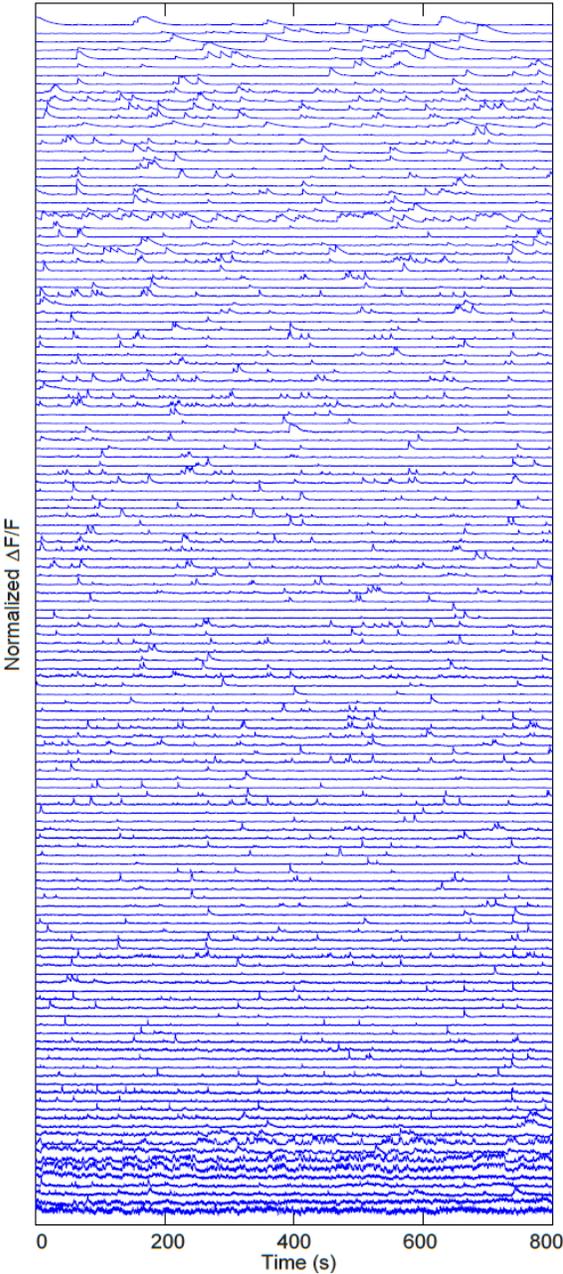
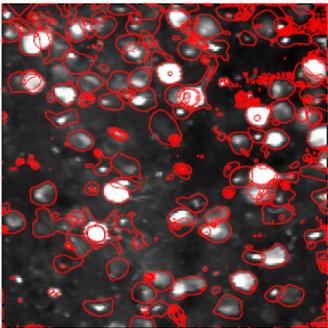
- (A) CNMF extracted spatial contour for 242 spatial components from a single plane recording of mouse V1 at depth of 500 μm from pial surface. Various spatial shapes of the spatial components including the doughnut shape and immediate perisomatic dendritic processes can be extracted.
- (B) The spatial contour for eight representative spatial components. The pixel weight for each spatial component is normalized for display purpose.
- (C)–(J) Normalized CNMF extracted temporal $\Delta F/F$ traces for the spatial components shown in (B). The right panel shows the zoom-in view of the region in the black dashed box in the left panel. The CNMF can accurately recover both the large signals with relatively long calcium decay times and small signals with short calcium decay times. The red traces show the clean CNMF extracted signals, and the blue traces show those including the residue.

Supplemental Figure S3 | Comparison of CNMF and ICA performance

A CNMF



B ICA

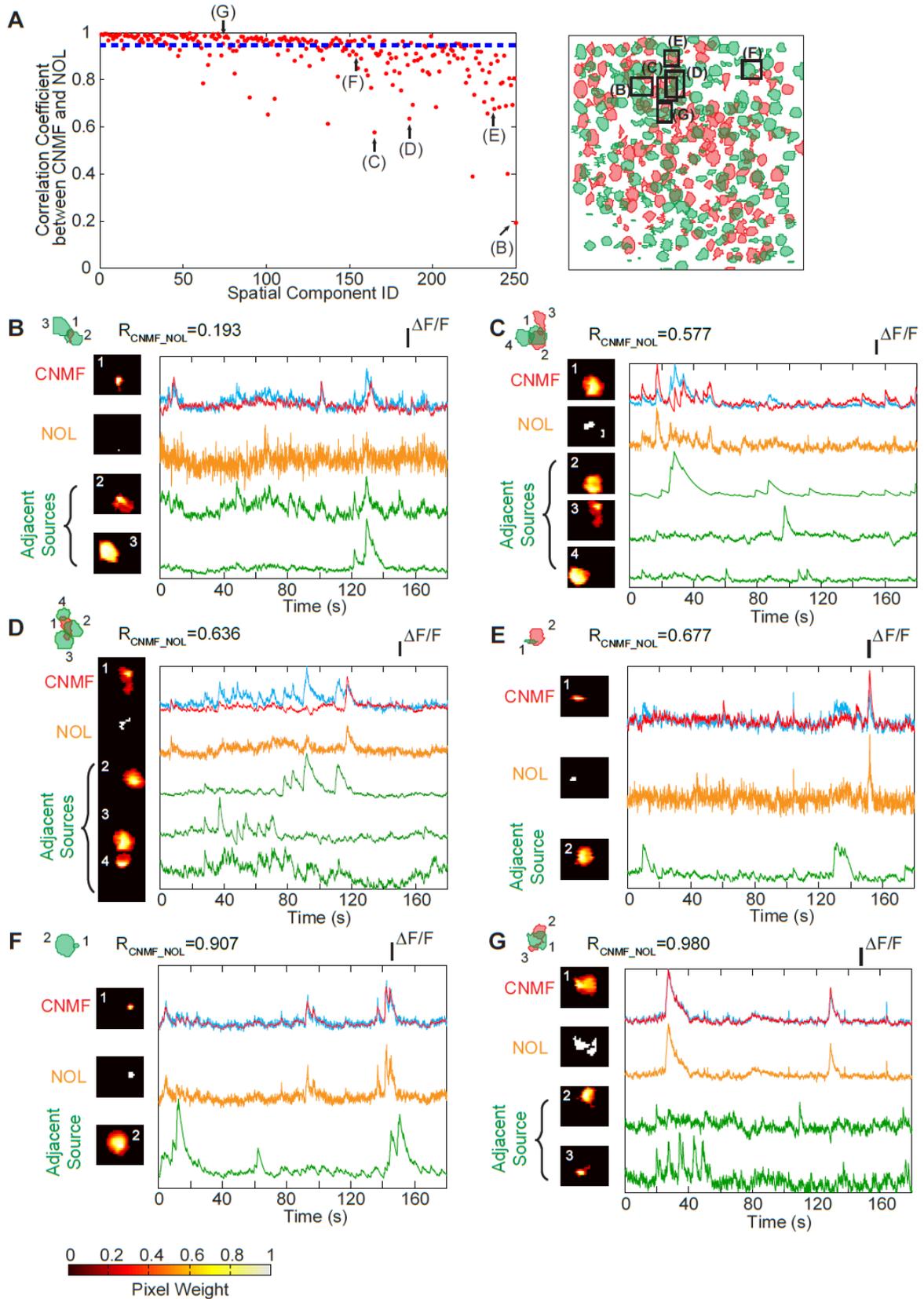


Comparison of CNMF and ICA performance on signal extraction from single-plane population calcium imaging. The data are from a single plane recording of mouse V1 at depth of 500 μm from the pial surface. CNMF finds more components with neuronal spatial profiles and typical calcium transients (with fast leading edges, and characteristic longer decays) compared to ICA.

(A) Spatial component contours and calcium signals extracted by CNMF. Top panel, spatial component contours overlaid on the temporal standard deviation image of a single plane recording of mouse V1 at depth of 500 μm from the pial surface. Bottom panel, normalized $\Delta F/F$ traces. Clean, cell like features are extracted, and the great majority of traces have signals with high SNR and typical calcium dynamics.

(B) Same as (A), with spatial component contours and calcium signals extracted by ICA. The displayed spatial contours were spatially smoothed (Gaussian blurred), and thresholded to show only positive components. Even with processing, the spatial components are less cell-like, and a large number of extracted traces show poor SNR, with no features recognizable as typical activity traces. Scale bar, 50 μm .

Supplemental Figure S4 | Correlation between signals extracted from CNMF and NOL

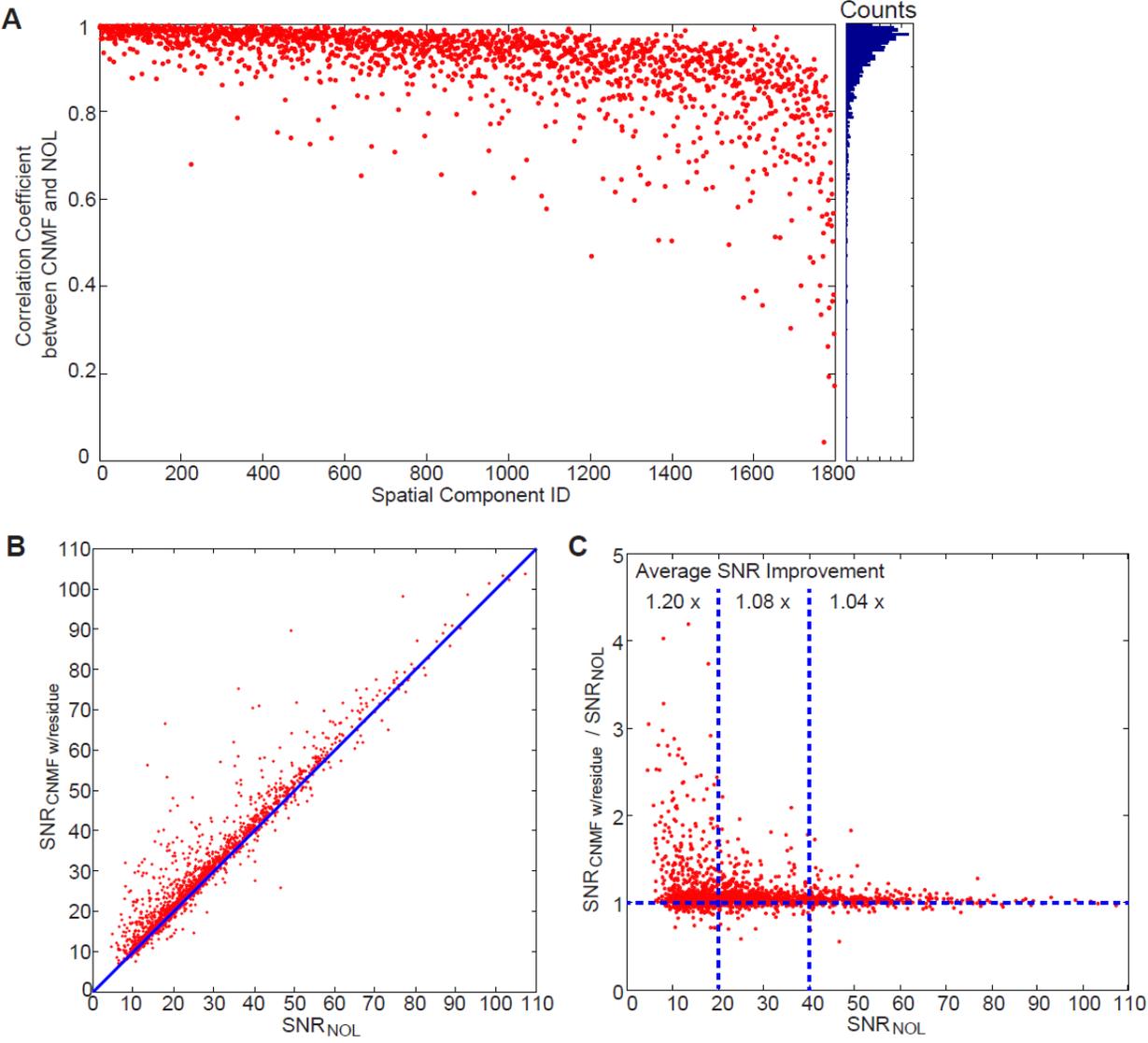


Correlation between signals extracted from CNMF and non-overlapped source regions (NOL). Majority of the spatial components show high correlations in the extracted temporal signal between CNMF and NOL. This is expected if the sample has some sparsity (some NOL region per spatial component), and the method cleanly extracts underlying dynamics. In the cases where there is low correlation, examination of underlying images shows why - the NOL signal is noisy due to either low pixel count, or contamination from nearby sources, or both; whereas the CNMF method is able to recover the signal while suppressing contamination, leading to a low correlation with that from NOL.

(A) Correlation coefficient between the $\Delta F/F$ extracted from CNMF and NOL spatial components (SCs) for a total of 250 spatial components, for the data in Figs. 3 and 4. The blue dashed line indicates the median of the correlation coefficients, and the contours for these spatial components are shown on the right. This particular figure panel is included only for convenience, and contains items from Fig. 5(A), and the bottom panel in Fig. 3(E).

(B)– (G) The spatial components with various correlation coefficients between CNMF and NOL are studied. For each case, the spatial component to be studied (labelled as 1) and its adjacent spatial components (potential contamination sources) are indicated in (A), and also re-plotted on top of the figure. To better evaluate the correlation between the temporal signal extracted by CNMF and NOL, both signals are plotted (CNMF in red, NOL in orange). A raw trace, generated from the CNMF identified spatial component, but with uniform pixel weighting, and without temporal demixing of overlapping sources, is plotted in cyan, superimposed onto the traces extracted with CNMF. The signals extracted from the adjacent spatial components (using CNMF) are plotted in green. The CNMF show a strong resistance towards signal contaminations from other sources, even in the case where the whole spatial component is nearly buried in other adjacent sources (i.e. example shown in B). Fluorescent traces are independently scaled for display convenience. The scale bar of $\Delta F/F$ is 0.39, 0.21, 0.59 for SC 1-3 in (B); 0.53, 0.83, 0.32, 0.57 for SC 1-4 in (C); 0.32, 0.52, 0.53, 0.19 for SC 1-4 in (D); 0.29, 0.22 for SC 1-2 in (E); 0.77, 0.31 for SC 1-2 in (F); 0.48, 0.17, 0.13 for SC 1-3 in (G).

Supplemental Figure S5 | Pooled statistics on CNMF performance over 10 different dual plane imaging sessions across 5 mice



Pooled statistics on CNMF performance over 10 different dual plane imaging sessions across 5 mice. Majority of the spatial components show high correlations in the extracted temporal signal between CNMF and NOL, and a higher SNR in CNMF than NOL. In those low correlation cases, the NOL signal is noisy due to a low pixel count, or contamination from nearby sources, or both; whereas the CNMF method is able to recover the signal while suppressing contamination. The signals in these cases thus receive a higher SNR improvement by using the CNMF algorithm.

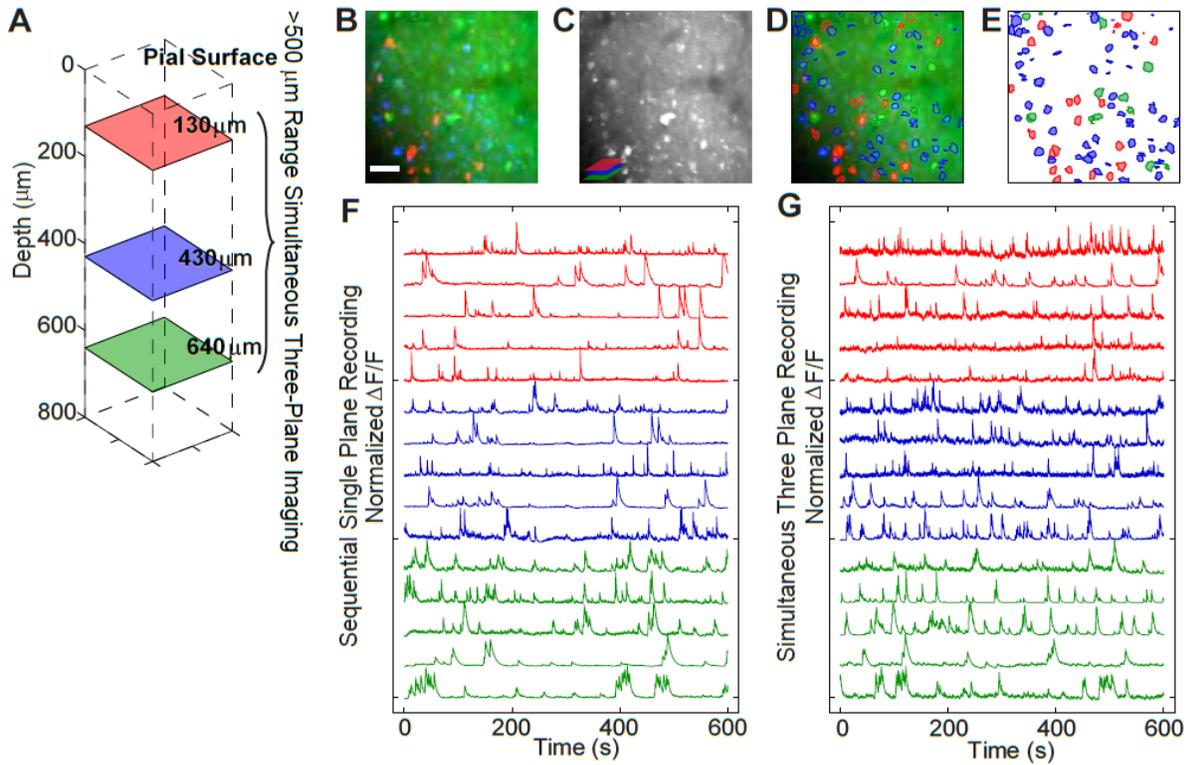
(A) Correlation coefficient between the $\Delta F/F$ extracted from CNMF and NOL for a total of ~1800 spatial components in the dual plane imaging. The histogram of the correlation coefficients are plotted on the right. The median and average of the correlation coefficient is 0.94 and 0.91 respectively. 29% of the spatial components have correlations below 0.9, which would benefit from a high performance demixing strategy, like CNMF. The spatial component ID is sorted by the SNR of their temporal signal extracted by NOL (from high to low). In general, the lower the SNR is, the lower the correlation coefficient.

(B) Signal to noise ratio (SNR) comparison between the $\Delta F/F$ extracted from CNMF (with residue) and the $\Delta F/F$ extracted from NOL for spatial components in (A). The average SNR from CNMF (with residue) is 12% higher than that from NOL.

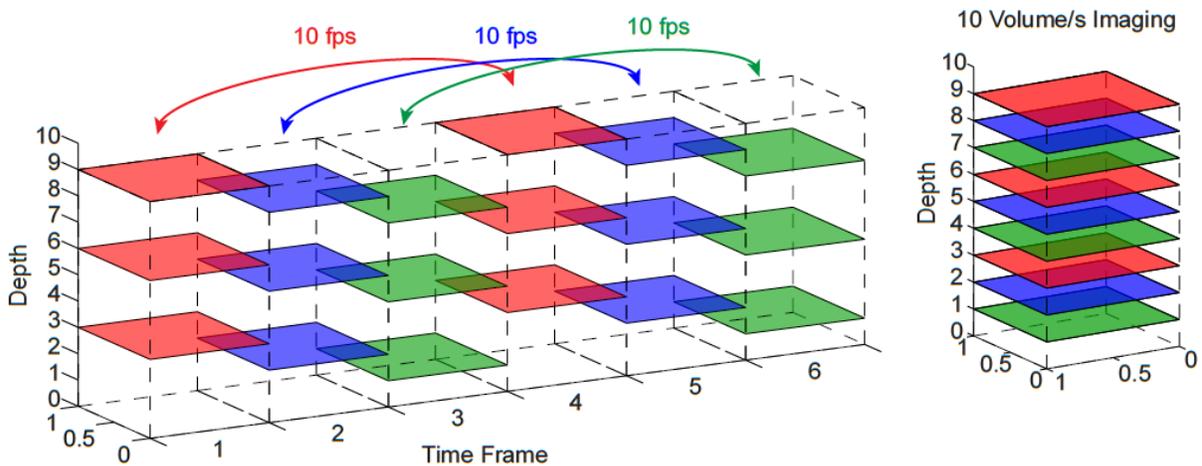
(C) Ratio of the SNR of the CNMF extracted signal (with residue) to the NOL signal, plotted against the SNR of the NOL extracted signal. This shows the relative gain in SNR from CNMF extracted sources as a function of the initial NOL SNR. In general, the extracted signals from sources with lower initial SNR using NOL slightly increase their SNR after being extracted with the CNMF algorithm.

Supplemental Figure S6 | Towards fast volumetric imaging: simultaneous axial three plane imaging extending the axial plane separation over 500 μm , and time-multiplexed strategy for SLM-based volumetric imaging

Three-Axial-Plane Imaging with $>500 \mu\text{m}$ Axial Separation



H Scheme of Time Multiplexed Volumetric Imaging



Towards high speed volumetric imaging by increasing the number and total axial range of simultaneously imaged planes, and utilizing an SLM time-multiplexing scheme.

- (A)– (G) Simultaneous three axial plane *in-vivo* functional imaging of mouse V1, at depth of 130 μm , 430 μm and 640 μm from pial surface, extending a simultaneous imaging range over 500 μm , with 10 fps imaging speed. The intermediate plane is set to be the same as the nominal focal plane where the zero order beam rests. In this additional experiment, the total performance and number of spatial components was limited by poor expression of the indicator in the mouse (limited number of cells expressing strongly, combined with high non-specific background fluorescence). Nonetheless, the overall method had little trouble extracting clean spatial source components.
- (A) Schematics of the simultaneous three-plane imaging extending the axial separation of the planes to over 500 μm .
- (B) Overlaid temporal standard deviation image of the sequential single plane recording of mouse V1 at depth of 130 μm (red), 430 μm (blue) and 640 μm (green) from pial surface of a mouse V1. Scale bar, 50 μm .
- (C) Temporal standard deviation imaging of the simultaneous three-plane recording of the three planes.
- (D) Overlaid spatial component contours from the three plane on (B).
- (E) Extracted spatial component contours.
- (F) Representative extracted $\Delta F/F$ traces of the selected spatial components from the three planes (red, 130 μm plane; blue, 430 μm plane; green, 640 μm plane), from the sequential single plane recording.
- (G) Extracted $\Delta F/F$ traces of the same spatial components shown in (N), from the simultaneous three-plane recording.
- (H) Illustration of a fast volumetric imaging scheme using time-multiplexed three-plane imaging, enabled by the high SLM switching speed between different holograms (> 300 Hz). In this example, the normal frame rate is 30 fps. With time-multiplexing between different sets of three-plane imaging, the volumetric imaging rate is 10 vol/s.

Supplemental Table S1. Zernike polynomials and coefficients for defocus, first-order spherical aberration and second-order spherical aberration

Defocus	
Zernike polynomials	$Z_2^0(u, v) = \sqrt{3} [2(u^2 + v^2) - 1]$
Zernike coefficients	$C_2^0(z) = \frac{nkz \sin^2 \alpha}{8\pi\sqrt{3}} \left(1 + \frac{1}{4} \sin^2 \alpha + \frac{9}{80} \sin^4 \alpha + \frac{1}{16} \sin^6 \alpha + \dots \right)$
First-order spherical aberration	
Zernike polynomials	$Z_4^0(u, v) = \sqrt{5} [6(u^2 + v^2)^2 - 6(u^2 + v^2) + 1]$
Zernike coefficients	$C_4^0(z) = \frac{nkz \sin^4 \alpha}{96\pi\sqrt{5}} \left(1 + \frac{3}{4} \sin^2 \alpha + \frac{15}{18} \sin^4 \alpha + \dots \right)$
Second-order spherical aberration	
Zernike polynomials	$Z_6^0(u, v) = \sqrt{7} [20(u^2 + v^2)^3 - 30(u^2 + v^2)^2 + 12(u^2 + v^2) - 1]$
Zernike coefficients	$C_6^0(z) = \frac{nkz \sin^6 \alpha}{640\pi\sqrt{7}} \left(1 + \frac{5}{4} \sin^2 \alpha + \dots \right)$
n , refractive index of media between the objective and sample; k , the wavenumber; z , the axial shift; u, v , coordinates on the SLM phase mask; $n \sin \alpha$, the NA of the objective.	

SLM Beam Steering

The overall system design of the SLM based two-photon microscope has been described in the main manuscript. In the following *Supplement*, we describe in depth further performance details of SLM-based beam steering, and the various factors that need to be considered when designing an SLM microscope.

Maximum lateral beam steering

The maximal lateral displacement of the incoming beam is controlled by the grating equation (Palmer, 2014):

$$m\lambda = \Lambda \cdot \sin(\theta), \quad (\text{S1})$$

where m is the diffraction order, λ the wavelength of the light, Λ , the spatial period of the feature, and θ the angle of deflection (from SLM). We consider only the primary order, $m = 1$, and use the imaging wavelength, $\lambda=940$ nm. For a pixelated device, the smallest periodic feature would consist of two pixels, each of which is $15 \mu\text{m}$ on our SLM. Thus $\Lambda_{\min} = 30\mu\text{m}$. We can relate the maximum SLM induced deflection angle to the maximum lateral displacement from the center of the FOV in the sample plane, r_{\max} , as follows:

$$r_{\max} = f_{\text{obj}} \cdot \tan(\theta_{\text{obj}}^{\max}) \quad (\text{S2})$$

where f_{obj} is focal length of the objective, and $\theta_{\text{obj}}^{\max}$ is maximum SLM induced deflection angle on the sample plane. We can relate this to the physical parameters of the SLM, intermediate optics, and objective through the following relations:

$$\theta_{\text{obj}}^{\max} = \theta_{\text{SLM}}^{\max} \cdot \left[\frac{1}{M_{\text{SLM}}} \right] \cdot \left[\frac{1}{M_{\text{scope}}} \right] \quad (\text{S3})$$

where $\theta_{\text{SLM}}^{\max}$ is the maximum SLM induced deflection angle at the SLM surface, M_{SLM} is the magnification of the 4f system right after the SLM, and M_{scope} is the magnification of the 4f system created by the scan lens and tube lens.

For our system, with a 25X objective, this corresponds to:

$$\begin{aligned}
 f_{obj} &= f_{tube\ lens} / M_{obj} = 180 \text{ mm} / 25 = 7.2 \text{ mm} \\
 M_{SLM} &= 100 / 400 = 0.25, \quad M_{scope} = 180 / 50 = 3.6 \\
 \theta_{SLM}^{max} &= \sin^{-1} \left(\frac{\lambda}{\Lambda} \right) = \sin^{-1} \left(\frac{940 \text{ nm}}{30 \text{ } \mu\text{m}} \right) = 1.8^\circ \\
 r_{max} &= 252 \text{ } \mu\text{m}
 \end{aligned}$$

For single point targeting, the *efficiency* of redirection depends on the number of phase levels available on the SLM. We measured our device, and found ~80 usable phase levels between 0 and 2π (Fig. S1 A). With the simplifying assumption that these levels are uniformly spaced between 0 and 2π , the first order diffractive efficiency of such a device (blazed grating) is given by:

$$\eta_1 = \left[\frac{\sin(\pi / N_p)}{\pi / N_p} \right]^2 \tag{S4}$$

where N_p is the total number of levels in each 2π phase ramp in the phase hologram (O’Shea et al., 2003). This implies very high efficiencies for holograms with large features (each 2π phase ramp spanning many pixels), which result in small positional displacements in the sample (Fourier) plane. For larger sample plane displacements, the required phase features get smaller and smaller (each 2π phase ramp spanning fewer and fewer pixels), and the finite sampling from the pixelated SLM reduces the available phase levels. In the extreme case, there is aliasing, which strongly degrades the performance – we avoid that limit for all holograms used in this paper. As a practical matter, we consider a minimum of four pixels across a 2π phase ramp to be the limit of efficient power redirection. The deflection vs. theoretical efficiency curve is shown in Fig. S1 D, along with the measured relative SLM power deflection efficiency on the sample (Fig. S1 C, D). For pixelated SLMs with quantized phase levels, the theoretical scalar beam redirection efficiency is reduced for deflections with a non-integral number of pixels per phase ramp, but this effect is small (Engstrom et al., 2008). At r_{max} , the maximum possible lateral deflection, the pattern necessarily has only two phase levels, which lowers the maximum diffraction efficiency, and additionally, results in a symmetric phase pattern. Under these conditions, the phase “grating” is not blazed, but instead splits the beam equally to the +1 and -1

orders – that is there are two beamlets spots, mirrored across the nominal center of the FOV, along with the residual zero order beam. This is not a problem for multi-region lateral imaging, where such a split can be advantageous. However, if the desired effect is to have a single targeted point on the sample, it is best to restrict the lateral targeting to regions that can be addressed by phase gratings that span at least four pixels (levels). With four pixels blaze and two photon excitation, the excitation ratio of the +1 to -1 order is $\sim 20:1$; for five pixels, $\sim 50:1$; and for eight pixels, $\sim 350:1$.

Maximum axial beam steering

Similar relationships hold for axial displacement. Here an SLM is used to impart a spherical curvature to the wavefront that shifts the focus before or beyond the nominal focal plane. For simplicity, we will only consider primary defocus here, neglecting higher order aberration terms. For a phase wrapped device, this would correspond to a Fresnel phase profile (O’Shea et al., 2003):

$$r_p^2 = 2pf\lambda + p^2\lambda^2 \xrightarrow{p\lambda \ll f} r_p^2 = 2pf\lambda \quad (\text{S5})$$

with r_p the radius, p , the Fresnel zone order, f , the focal length, and λ the wavelength. Each successive zone corresponds to adding one wave of defocus at that point on the pupil, which also corresponds to a 2π phase change. For large p , the relative spacing between zones is $\Delta r_p = f\lambda / r_p$. The “strongest” lens allowed without aliasing on an actual device would correspond to a two pixel spacing on the outermost Fresnel zone. In our chosen SLM, with the pixel pitch d being $15 \mu\text{m}$, and the number of pixels on the SLM across the shortest dimension N being 512, the strongest focal length corresponds to

$$\|f\| = \frac{\Delta r_p \cdot r_p}{\lambda} = \frac{2d \cdot \frac{1}{2}Nd}{\lambda} = d^2 \cdot \frac{N}{\lambda} = 123 \text{ mm} \quad (\text{S6})$$

We note that the sign of f can be made positive or negative simply by reversing the phase. With this magnitude for the focal length on the SLM, and our intermediate optical scaling onto the objective back aperture (effective excitation NA ~ 0.45 for the 25X objective), this lens power corresponds to axial displacements of $\pm 520 \mu\text{m}$ under the objective, calculated using thin lens

formula. Any larger displacement will necessarily result in aliasing on the SLM, which both reduces diffraction efficiency, as well as creates secondary foci.

For larger axial displacements without aliasing, the SLM would need more pixels. This could be done by shrinking the pixels, growing the size of the SLM, or both. We state our general preference with current LCOS devices would be to have more pixels on a larger SLM rather than smaller pixels on same device – smaller pixels typically have lower fill factors, and larger pixels may have increased power handling, as the overall power density on the physical device is decreased with increased SLM area.

Similar to the lateral deflection, the efficiency of the SLM drops with an increased focus shift. In designing our system, we set a goal of being able to holographically deflect light over a span of ~500 μm axially, while maintaining clear subcellular resolution and high two-photon efficiency. Our measured efficiency curve is slightly asymmetric (Fig. S1 B), but shows that we can project the beam from 200 μm beyond (deeper) the focal plane of the objective, to 300 μm above (shallower) the objective's focal plane while maintaining strong two-photon excitation efficiency. We rely on this curve throughout this study, to mechanically set the objective's natural focal plane to a specific position for optimal power delivery to the chosen targeted planes in simultaneous imaging, and to compute the power necessary for good signal at various multiplane combinations; this is further discussed in the following sections.

SLM pixel number, pixel size, and fill factor

The overall efficiency of this method can be improved by better SLMs; that is devices that have higher fill factors, increased pixel number, and increased phase modulation – all of which are the subject of active development. As already stated earlier, increased pixel counts increases the possible number of levels available across any feature, and will increase the diffraction efficiency. For arbitrary patterns of excitation, it is not possible to predict how many overall pixels one would need on the device to achieve this, but for the range of simple lateral or axial deflections presented in this paper, a device with ~1500 pixels across would keep the overall module power efficiency above 80% for the current fill factor, by providing ~4 pixels per phase ramp. Besides pixel counts, pixel size also plays a role in the overall SLM performance. Large pixels have the added benefit of increased power handling, and generally higher fill factors, because the electronics and insulating barriers between neighboring pixels are somewhat fixed in

size, whereas active areas are not. Larger pixels will result in larger phase mask features, which reduces the maximum angles of deflection, but this is easily compensated with altering the magnification of the post-SLM telescope system, which is necessary anyway to properly map the SLM to the back focal plane of the objective.

The fill factor of the SLM is defined as the ratio between the area of all the active elements of the SLM and the total area of the SLM. It is important for two reasons. The first is straightforward – the overall power transmission to the diffracted beam is directly proportional to the fill factor. The pixel fill factor of the SLM used here is 82%, and thus even with “perfect” holograms, the maximum power throughput is reduced compared to a simple mirror. The second reason is more subtle. For a given diffraction pattern, having fill factors less than one broadens the far-field envelope that governs the relative efficiencies between different diffraction orders. As a result, power is transferred from the desired first order into higher orders, which is not available for excitation and can generate spurious background signals (Arrizon et al., 1999). For additional detailed discussion on the limits of pixelated spatial light modulators in neuroscience, see Golan et al., and references therein (Golan et al., 2009).

Steering dependence on objective

The focal length of the objective is given by $f_{obj} = f_{tube\ lens} / M_{obj}$. Together with the relationship between the beam lateral displacement and the objective focal length given in Eq. S2, it is apparent that the lateral extent is inversely proportional to the magnification. Thus a reduction of objective magnification by a factor of two doubles the lateral displacement of the beam on the sample for the same grating pattern on the SLM. The lateral displacement is independent of the NA of the objective, and the effective NA of the illumination beam, and depends only on the tilt angle of the wavefront with respect to the optical axis at the pupil of the objective. However, the effective NA can affect the relative efficiency of two photon excitation through chromatic effects, for larger angular displacements, as described in later sections.

For axial displacements, with geometric ray optics, the amount of defocus is derived as:

$$dz = \frac{2 \cdot W_z \cdot \lambda \cdot n}{NA^2} \quad (S7)$$

where, dz is axial displacement from focal plane, n , the refractive index, W_z , the number of waves of defocus at the edge of the pupil, λ , the wavelength, and NA , the numerical aperture. As written, this equation would assume that the image of the SLM exactly matches the size of the pupil (back aperture) of the objective. In our system, this is not the case, and the size of the SLM as imaged onto the back focal plane of the objective is smaller than the full back aperture, so the back aperture is only partially filled, and $NA_{eff} \leq NA_{obj}$, which constrains the maximum excitation NA . We recast this equation with the number of waves across the SLM, leading to:

$$dz = \frac{2 \cdot W_{zSLM} \cdot \lambda \cdot n}{NA^2} \cdot \left(\frac{NA}{NA_{eff}} \right)^2 = \frac{2 \cdot W_{zSLM} \cdot \lambda \cdot n}{NA_{eff}^2},$$

$$\frac{NA_{eff} = \frac{r_{S-SLM}}{f_{obj}} \cdot n}{\rightarrow} dz = \frac{2 \cdot W_{zSLM} \cdot \lambda \cdot n}{\left(\frac{r_{S-SLM}}{f_{obj}} \cdot n \right)^2}, \quad (S8)$$

$$\frac{M_{obj} = \frac{f_{tube}}{f_{obj}}}{\rightarrow} dz = \frac{2 \cdot W_{zSLM} \cdot \lambda \cdot n}{\left(M_{obj} \cdot \frac{r_{S-SLM}}{f_{tube}} \cdot n \right)^2} = \frac{2 \cdot W_{zSLM} \cdot \lambda}{\left(M_{obj} \cdot \frac{r_{S-SLM}}{f_{tube}} \right)^2 \cdot n}$$

where W_{zSLM} is the number of waves of defocus at the edge of the SLM surface, r_{S-SLM} is the radius of the SLM as imaged on the pupil, f_{tube} , f_{obj} , and M_{obj} as defined earlier in the supplemental information. Thus for fixed W_{zSLM} , on a given microscope, changes in magnification of the objective scale the axial displacement by SLM by:

$$\frac{dz_1}{dz_2} = \left(\frac{M_{obj2}}{M_{obj1}} \right)^2 \quad (S9)$$

Chromatic Effects

SLMs are diffractive optics, so they necessarily are chromatic devices – that is beam redirection depends on wavelength (see Eq. S1). Thus any non-monochromatic source (e.g. a femtosecond pulsed laser) will exhibit chromatic effects. The relationship connecting bandwidth to pulse width is $\Delta\nu\Delta t \geq K$, with $K=0.315$ for the hyperbolic secant-squared shaped pulses from most Ti-

Sapphire lasers (Hirlimann, 2005). Our laser produced transform-limited pulses of 140 fs, which means the FWHM of the nominally 940 nm beam is 6.5 nm. Considering the blue and red edge of this spectral profile, for displacements in the x-y plane, there is a shift in the lateral position of the spot across the spectrum. With a 25X objective, and an SLM induced lateral displacement of 150 μm from the origin for the central beam (which is the largest lateral displacements used in this paper), we find the spectral FWHM “colors” to be deflected $\pm 0.6 \mu\text{m}$ from the target position (calculated using Eq. S1 and S2). These displacements are comparable in size with the FWHM of the monochromatic PSF, which both blurs the image and reduces the peak focused intensity at such displacements. The reduced intensity can be partially compensated with appropriate weighting of the targets during hologram calculation (Fig. S1 E and F), but the “blur” cannot. With shorter pulses, performance fall off would be more severe, and without additional correction elements, using this splitting method with pulse widths shorter than 100 fs is not recommended. One should also note that without additional correction, the relative drop in two photon efficiency due to lateral color increases with increasing effective NA for the same lateral displacement.

This chromatic dispersion can be reduced with the incorporation of a custom shaped high dispersion optical element, as demonstrated in Katona (Katona et al., 2012), or a combined diffractive/refractive device similar to that used in Abrahamsson (Abrahamsson et al., 2013), to improve the performance over a wider lateral range, and for shorter pulses. Nonetheless, even without any effort to reduce chromatic dispersion, we are easily able to extend the FOV a meaningful amount, rapidly and flexibly, as demonstrated in Fig. 2.

For axial displacements on our system, the effect of chromatic dispersion on performance is markedly less. The focal length variation with wavelength - axial color can be expressed as

$$\frac{\Delta dz}{dz_0} = \frac{\Delta \lambda}{\lambda_0}, \text{ following Eq. S8. This corresponds to an axial shift between the FWHM points of}$$

$\sim 1.4 \mu\text{m}$ for a $-200 \mu\text{m}$ displacement (longer focal length), and $\sim 2.1 \mu\text{m}$ for a $+300 \mu\text{m}$ displacement (shorter focal length). We fill the objective back aperture to have an effective NA of 0.45, and with these settings, the FWHM of the nominal monochromatic two-photon axial PSF is $\sim 6.5 \mu\text{m}$ (Zipfel et al., 2003), so the chromatic axial displacement is only a small fraction of the PSF for the entire span of axial separations demonstrated here. This has a negligible effect

on the axial resolution, and power delivery for somatic imaging. If finer axial resolution is required, which requires larger excitation NA, the axial range with high performance is correspondingly smaller. For a given SLM setting, the relative displacement the SLM produces, and the resultant axial PSF, both scale as NA^2 . For example, at an excitation NA of 0.9, 125 μm of axial separation are possible with the same relative performance. For our chosen task requirement, however, fast multiplane calcium imaging with single cell resolution, the performance is excellent over 500 μm with an excitation NA of 0.45.

In the above analysis, we ignored any possible variations in efficiency from shifts in the optimal phase depth as a function of wavelength. These variations would occur because the bandwidth is finite, and there is a wavelength dependence to the retardation in the liquid crystal. These effects are likely to be very small, because the liquid crystals used in our SLM have only very minor variations in refractive index as a function of wavelength around our chosen center.

For axial focusing, the relative efficiency of a diffractive lens as a function of wavelength can be expressed as (O'Shea et al., 2003):

$$\eta = \frac{\sin^2[\pi(\alpha-m)]}{[\pi(\alpha-m)]^2}, \text{ with } \alpha = \frac{\lambda_0}{\lambda} \left[\frac{n(\lambda)-1}{n(\lambda_0)-1} \right], \text{ and } m \text{ the diffraction order}$$

For the bandwidth of our laser, the refractive index, $n(\lambda)$, is essentially constant, hence

$$\alpha = \frac{\lambda_0}{\lambda} \Rightarrow [0.9965 \leq \alpha \leq 1.0035] \text{ which implies } \eta \rightarrow 1 \text{ for lenses operating in the first order.}$$

This implies that we can ignore the finite wavelength spread of the beam when considering the diffraction efficiency of the SLM in axial focusing.

Adaptive optics

The overall excitation efficiency can be increased by combining adaptive optics methods with our targeting. SLMs are a natural choice for corrections of both system and sample aberrations (Ji et al., 2012). We implement a few basic corrections in this paper, namely higher order spherical aberrations from defocus itself. We originally expected to require additional corrections to perform deeper multilayer imaging, but it did not prove to be necessary for high SNR recordings even below 600 μm in the cortex. Nonetheless, minimizing the power on sample while maximizing signal can only increase the quality of imaging, and minimize any heat induced perturbations, so it is an area of ongoing effort, in our laboratory, and others.

Power delivery to each axial layer

The mouse brain is highly scattering (Dunn, 2014; Kobat et al., 2011), and there is an exponential loss of ballistic photons with increasing depth. The nominal intensity in the focus as a function of depth is given by $I_d = I_0 \exp\left[\frac{-D}{l_s}\right]$, where I_0 is the incident intensity at the surface of the brain, D , the depth, and l_s , the characteristic scattering length. For 940 nm, the scattering length is $\sim 185 \mu\text{m}$ (Horton et al., 2013). For intensities below the saturation limit, the two-photon fluorescence signal scales as the intensity squared. If we assume that the expression of the indicator is the same in the desired layers chosen for imaging, we can easily estimate the relative power that needs to be delivered to each layer to give equivalent fluorescence signal.

$$\begin{aligned} F_1 &\propto \sigma \left(I_{01} \cdot \exp\left[\frac{-d_1}{l_s}\right] \right)^2, \quad F_2 \propto \sigma \left(I_{02} \cdot \exp\left[\frac{-d_2}{l_s}\right] \right)^2, \\ F_1 \approx F_2 &\Rightarrow \left(I_{01} \cdot \exp\left[\frac{-d_1}{l_s}\right] \right)^2 = \left(I_{02} \cdot \exp\left[\frac{-d_2}{l_s}\right] \right)^2, \\ \frac{I_{01}}{I_{02}} &= \exp\left[\frac{d_1 - d_2}{l_s}\right] \end{aligned} \tag{S10}$$

For the experiment shown in Fig. 3, this ratio is as follows:

$$\begin{aligned} \frac{I_{01}}{I_{02}} &= \exp\left[\frac{170 - 500}{185}\right] = 0.168, \\ I_{02} &\approx 6 \cdot I_{01} \end{aligned}$$

The signal generated also depends critically on the overall efficiency of the SLM in redirecting the light to axial positions other than the designed focal plane of the microscope (Fig. S1 B). In this example, the microscope's natural focal plane is set to be $380 \mu\text{m}$, and the measured SLM efficiency for redirection is approximately the same for both planes ($170 \mu\text{m}$ and $500 \mu\text{m}$). We thus compute the dual beam hologram to have an approximately six-fold increase in intensity directed to the deeper layer compared with the superficial layer, such that the collected fluorescence from each plane was approximately equal.

Hologram generation

To create a 3D beamlet pattern at the sample (a total of N beamlets, each with coordinates $[x_i, y_i, z_i]$, $i=1,2\dots N$), the phase mask on the SLM, $\phi(u, v)$, can be expressed as follows:

$$\phi(u, v) = \text{phase} \left\{ \sum_{i=1}^N A_i e^{2\pi j \{x_i u + y_i v + [Z_2^0(u, v)C_2^0(z_i) + Z_4^0(u, v)C_4^0(z_i) + Z_6^0(u, v)C_6^0(z_i)]\}} \right\} \quad (\text{S11})$$

A_i is the electrical field weighting factor for the individual beamlet. $Z_m^0(u, v)$ and $C_m^0(z_i)$ are the Zernike polynomials and Zernike coefficients, respectively, which fulfills the defocusing functionality and compensates some of the higher order spherical aberration due to defocusing. The expressions of $Z_m^0(u, v)$ and $C_m^0(z_i)$ are detailed in Supplemental Table S1 (Anselmi et al., 2011).

It is important to optimize the A_i coefficients in Eq. S11 so as to control the power delivered to each beamlet. In general, it is desired to configure A_i such that the fluorescence intensity coming from each beamlet is similar. It is important to note that this coefficient is not simply the desired power ratio – rather it affects the interference between the different light fields, which in turn depends on the relative phase displacements of the beams. Operationally, we can adjust the parameter empirically to achieve similar fluorescent signals from different imaging planes. Alternatively, for axial defocusing, based on the defocusing length, we rely on the efficiency curve shown in Fig. S1 B as well as the tissue scattering length (see the above section), and perform numerical beam propagation of the electric field (Schmidt, 2010) to determine A_i ; for lateral displacement, we mainly rely on the efficient contour shown in Fig. S1 C. Fig. S1 E shows an example of lateral displacement where we set A_i to be uniformed across different points, and as a result, the fluorescence drops off from the center field of view to the edge. After applying a proper weighting of A_i based on Fig. S1 C, a more uniform fluorescence across the field of view is obtained in Fig. S1 F. Following these guidelines, we achieve high performance imaging such as those demonstrated in Fig. 1~3, 6~7 and Fig. S6.

In our study, we typically computed the phase mask with a direct approach using Eq. S11, and found this to be sufficient for all experiments because of the small number of beamlets generated. For experiments that would require a very large number of beamlets, or holographic area rather than point stimulation, iterative adaptive approaches that utilize additional constraints

such as spatial and phase quantization, as outlined in Nikolenko et al. (Nikolenko et al., 2008), or a weighted Gerchberg-Saxton approach typically yield more uniformity across the field.

Regardless of the approach used for hologram computation, some care needs to be used in choosing targets. For patterns with a very high degree of symmetry, linear superposition of the individual phase profiles can result on overall phase profiles with a low number of effective phase steps. For instance, if one selects axial defocus at $\pm 100 \mu\text{m}$, the resulting phase profile will have only two phase levels, and no “blaze”. In this configuration, $\sim 20\%$ of the power remains in the zero order, which may introduce artifacts in imaging if it is not considered. As a result, we typically strive for some asymmetry in the beam placement with respect to the origin for even numbers of desired beamlets. In instances where highly symmetric patterns were used (the dual lateral plane imaging), a small zero order block was placed in the intermediate image plane of the post SLM telescope (small metallic beam block on a thin pellicle).

SLM switching speed

The SLM used in this experiment has a high switching speed between different hologram patterns. It can be controlled via low latency PCIe bus transfers, which increases speed and gives deterministic latency for pattern changes. This device is also optimized for fast hologram transitions by increasing the backplane drive voltage, utilizing a more than 2π phase stroke and the optimized intermediate holograms during switching. These device enhancements dramatically increase the effective switching speed of the device, and allows for beam retargeting in $< 3\text{ms}$. Fig. 7J-K highlight the performance. We show that we can fully transition between the two sets of planes in 3 ms, with usable fluorescence produced in the new set of planes within 2 ms. Transitioning the system to resonant galvanometers should allow acquisition at $\sim 30 \text{ Hz}$, with a target of three interleaved images, each with three simultaneously illuminated planes (schematics shown in Fig. S6 H). The fast transition times insure that we can have fast interframe switching while maximizing duty cycle. This would allow for 9 axial sections, with an overall volume rate of 10 Hz, over large regions of the cortex, and at significant depth. The imaged volume can also be increased by simply scaling up the number of independent beams entering the SLM modules, and incorporating the spatial and temporal multiplexing strategies used in (Cheng et al., 2011; Stirman et al., 2014). In these cases, the addressable volume or speed increase scales linearly with the number of beams entering the SLM module(s).

Beam configurations for multiplane imaging

Since the SLM is flexible in its ability to address arbitrary sub-regions of the FOV, we can perform various experimental paradigms (data not shown) beyond those shown in Fig. 1B and C. For instance, one could double the effective frame rate of a system by creating two laterally displaced beams, with an angular spread one half of the FOV, in the direction orthogonal to the fast axis of the galvanometers. By scanning the galvanometers over the middle 50% of the image, the displaced beams still illuminate the entire FOV, though the total number of lines scanned is halved; this effectively doubles the overall frame rate. Alternately, two or more small sub-regions enclosed within a single larger FOV can be scanned simultaneously. Under all of these paradigms, the splitting allows for increases in the effective frame rate, while keeping the dwell time per pixel of each region higher than what would have been possible if the regions were sequentially scanned, which increases the overall signal collected from each region.

Source Location Initialization in CNMF Algorithm

The CNMF algorithm has been described in the main manuscript. Here we further describe how the locations of the sources are initialized in this algorithm. By default the algorithm's spatial component detection algorithm initializes source locations by finding regions in the image with high local variance. It can instead be initialized with a list of spatial component centroids. This list can be generated by any external method the user chooses. The default code will bias detection towards active (time fluctuating) signals, and/or bright regions. As such, it will not detect cells that are inactive and dim, and the algorithm is not meant for determining the total number of sources, active and inactive, in a region. Nonetheless, we have found that the method is sufficiently sensitive to capture more active cells (with higher SNR) than other automated methods (see Fig. S3). In cases where one may want to know every cell's location, regardless of activity, cells could be co-labeled with a different color fluorophore, and that image can be used to generate the spatial component list.

Supplemental References

Abrahamsson, S., Chen, J., Hajj, B., Stallinga, S., Katsov, A.Y., Wisniewski, J., Mizuguchi, G., Soule, P., Mueller, F., Darzacq, C.D., *et al.* (2013). Fast multicolor 3D imaging using aberration-corrected multifocus microscopy. *Nature Methods* *10*, 60-U80.

Anselmi, F., Ventalon, C., Begue, A., Ogden, D., and Emiliani, V. (2011). Three-dimensional imaging and photostimulation by remote-focusing and holographic light patterning. *Proceedings of the National Academy of Sciences of the United States of America* *108*, 19504-19509.

Arrizon, V., Carreon, E., and Testorf, M. (1999). Implementation of Fourier array illuminators using pixelated SLM: efficiency limitations. *Optics Communications* *160*, 207-213.

Cheng, A., Goncalves, J.T., Golshani, P., Arisaka, K., and Portera-Cailliau, C. (2011). Simultaneous two-photon calcium imaging at different depths with spatiotemporal multiplexing. *Nature Methods* *8*, 139-U158.

Dunn, A.K. (2014). Optical Properties of Neural Tissue. In *Optical Imaging of Neocortical Dynamics*, B. Weber, and F. Helmchen, eds., pp. 33-51.

Engstrom, D., Bengtsson, J., Eriksson, E., and Goksor, M. (2008). Improved beam steering accuracy of a single beam with a 1D phase-only spatial light modulator. *Optics Express* *16*, 18275-18287.

Golan, L., Reutsky, I., Farah, N., and Shoham, S. (2009). Design and characteristics of holographic neural photo-stimulation systems. *Journal of Neural Engineering* *6*.

Hirlimann, C. (2005). *Femtosecond Laser Pulses: Principles and Experiments*, Vol Advanced texts in physics, 2nd edn (Springer).

Horton, N.G., Wang, K., Kobat, D., Clark, C.G., Wise, F.W., Schaffer, C.B., and Xu, C. (2013). In vivo three-photon microscopy of subcortical structures within an intact mouse brain. *Nature Photonics* *7*, 205-209.

Ji, N., Sato, T.R., and Betzig, E. (2012). Characterization and adaptive optical correction of aberrations during in vivo imaging in the mouse cortex. *Proceedings of the National Academy of Sciences of the United States of America* *109*, 22-27.

Katona, G., Szalay, G., Maak, P., Kaszas, A., Veress, M., Hillier, D., Chiovini, B., Vizi, E.S., Roska, B., and Rozsa, B. (2012). Fast two-photon in vivo imaging with three-dimensional random-access scanning in large tissue volumes. *Nature Methods* *9*, 201-208.

Kobat, D., Horton, N.G., and Xu, C. (2011). In vivo two-photon microscopy to 1.6-mm depth in mouse cortex. *Journal of Biomedical Optics* *16*.

Nikolenko, V., Watson, B.O., Araya, R., Woodruff, A., Peterka, D.S., and Yuste, R. (2008). SLM microscopy: scanless two-photon imaging and photostimulation with spatial light modulators. *Frontiers in Neural Circuits* *2*.

O'Shea, D.C., Suleski, T.J., Kathman, A.D., and Prather, D.W. (2003). *Diffraction Optics: Design, Fabrication, and Test*, Vol TT62 (Bellingham, Washington, USA: SPIE Press).

Palmer, C. (2014). *Diffraction Grating Handbook*, 7th edn (THERMO RGL).

Schmidt, J.D. (2010). *Numerical Simulation of Optical Wave Propagation with Examples in MATLAB* (Bellingham, Washington 98227-0010 USA: SPIE Press).

Stirman, J.N., Smith, I.T., Kudenov, M.W., and Smith, S.L. (2014). Wide field-of-view, twin-region two-photon imaging across extended cortical networks.

Zipfel, W.R., Williams, R.M., and Webb, W.W. (2003). Nonlinear magic: Multiphoton microscopy in the biosciences. *Nature Biotechnology* *21*, 1369-1377.