An Optics-Free Ultra-Thin Time-Resolved Intraoperative Imager Using Upconverting Nanoparticles

Hossein Najafiaghdam
Ali Niknejad, Ed.
Mekhail Anwar, Ed.

Electrical Engineering and Computer Sciences
University of California, Berkeley

Technical Report No. UCB/EECS-2022-170
http://www2.eecs.berkeley.edu/Pubs/TechRpts/2022/EECS-2022-170.html

June 13, 2022

Acknowledgement

I would like to acknowledge Efthymios P. Papageorgiou for all the helpful tips and support he provided me throughout my studies at UC Berkeley.

An Optics-Free Ultra-Thin Time-Resolved Intraoperative Imager Using Upconverting Nanoparticles

By

Hossein Najafiaghdam

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering- Electrical Engineering and Computer Sciences

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Ali Niknejad, Chair
Professor Vladimir Stojanovic
Professor Mekhail Anwar
Professor Michel Maharbiz

Summer 2022

Abstract

An Optics-Free Ultra-Thin Time-Resolved Intraoperative Imager Using Upconverting Nanoparticles

by

Hossein Najafiaghdam

Doctor of Philosophy in Electrical Engineering and Computer Sciences

University of California, Berkeley

Professor Ali Niknejad and Mekhail Anwar

The successful outcome of cancer therapy and treatment depends on the ability to ensure no microscopic residual disease is left behind during the treatment. It is therefore critical to be able to visualize and identify any residual disease being left behind in the tumor bed during resection surgeries, preferably in real-time and in an intraoperative setting. For practical purposes, leaving behind any more than 200 cancerous cells in the tumor bed increases the chance of cancer recurrence -across all cancer types-, highlighting the importance of microscopic residual disease. Recent advancements in fluorescently-tagged targeted molecular probes and imaging agents have enabled a significantly enhanced selectivity in detecting cancer cells, allowing single cell detection using conventional fluorescent imaging techniques. However, these techniques have remained largely impractical in intraoperative settings due to the fact that they rely heavily on large and cumbersome instruments, including rigid optical filters (for color and wavelength selectivity) and focusing lenses to be able to resolve the image from their operating distance. The bulky and rigid optical lenses and filters, required to resolve the weak fluorescence signal from background, are challenging to miniaturize, and restrict the imager to a relatively far working distance from the tumor cells, reducing both the sensitivity and maneuverability within complex tumor cavities. These limitations also impose a minimum size and form factor restriction on these optical imagers, precluding a majority of them from being deemed practical in surgical settings, and particularly hard to maneuver in today's minimally invasive procedures.

Ideal intraoperative molecular imaging platforms will not only require an ultra-small form factor sensor with high sensitivity (<1000 cells), but also the ability to perform deep tissue imaging to enable extracting valuable information from not just the tumor bed, but also the lymph nodes that are often sitting a few millimeters below the resection sites. This imager should be able to resolve images without any optics (focusing lenses, wavelength filters) and be compatible with a molecular probe that has the ability to be excited with a deep-penetrating wavelength such as in the near infrared (NIR) range.

In this thesis, we shift the cumbersome optical requirements of fluorescence imaging into the time domain, using an optics-free micro-fabricated, time-resolved contact imaging array. Made possible through the synergistic integration of a custom high-speed integrated circuit imaging array and ultra-efficient upconverting nanoparticles (UCNPs), which are 2-3 orders of magnitude brighter than conventional lanthanide-doped UCNPs and have long (>100μs) phosphorescence

lifetimes, this work provides an integrated and standalone imaging platform using a 200μm-thin imager, which functions as a scalable molecular imaging skin, able to be integrated on any surgical instrument, for real-time visualization of tumor cells in surgical settings. This sensor has been designed to be synergistically compatible with UCNPs, and their excitation wavelengths of either 980 nm or 1550 nm, for a 2-photon or 3-photon absorption mechanism respectively.

This image sensor, measuring 4.8 mm by 2.3 mm, consists of an 80-by-36 pixel-array and includes custom-fabricated integrated micro-collimators to enhance resolution and sharpness when performing direct contact imaging, and by proxy obviating focusing lenses. To eliminate wavelength filters, a time-domain image resolution technique is implemented in which pulses of excitation light (980 nm or 1550 nm) with acquisition windows, taking advantage of the UCNP's long emission lifetime. To mitigate the challenge of the effect of the background generated by the pulsed excitation light, we have implemented a novel dual-photodiode pixel architecture where a secondary diode is used to measure the local background level and later used to remove it from the main signal. We have optimized area consumption by the two in-pixel photodiode and maintained a fill factor of 47% and implemented a pixel-level cancellation scheme using a one-time and initial non-linear curve fitting to fine-tune the ratios of the two photodiodes.

Using power levels compatible with *in vivo* use, we achieve cellular level detection and have validated the performance of this sensor using a 980 nm excitation with an intratumorally injected prostate tumor specimen in mouse and have been able to achieve a signal to background (SBR) level of 8 with an excitation power of 45 W/cm^2 using pulsed 980 nm light. The imager is also able to reduce the background generated by the excitation light to single digit mV level, nearing its noise floor level of 2.2 mV rms.

The imager developed in this thesis demonstrates the capability of optics-free CMOS imagers in being integrated in surgical settings and becoming miniaturized alternatives for large and cumbersome conventional fluorescent imagers.

To Shadow, my friends and my family.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Motivation

Treatment of cancer with surgery and radiation relies heavily on the ability to both visualize and quantify tumor cell distribution to direct where the treatment should be focused; however, this information is not presently available for microscopic tumor foci leaving physicians to empirically treat wide areas of tissue, causing significant toxicity, or risk leaving untreated tumor cells behind, worsening cancer outcomes. While gross (>1cm) tumor is detectable by eye and touch, microscopic residual disease (MRD) consisting of clusters 100s and 1000s of tumor cells morphologically similar to their normal tissue counterparts, cannot be readily visualized. The ability to precisely identify, localize and quantify the spread of the tumor cells during surgery is paramount to the successful outcome of the operation and, if necessary, post-operative radiation. Failure to properly resect and remove residual tumor around the periphery of tumor cavities dramatically increases the risk of cancer recurrence, for example in breast cancer doubling the risk of recurrence from 15% to 30% over a span of 10 years [1]. **Figure 1** demonstrates the stark increase in cancer recurrence rates for several types of cancers in cases where microscopic residual disease was left behind during surgery. A similar challenge occurs in prostate cancer where up to 50% of high-risk prostate cancer patients are found to have MRD post-operatively and must undergo 8 weeks of radiation and hormone therapy. Cancer recurrence is a serious challenge across

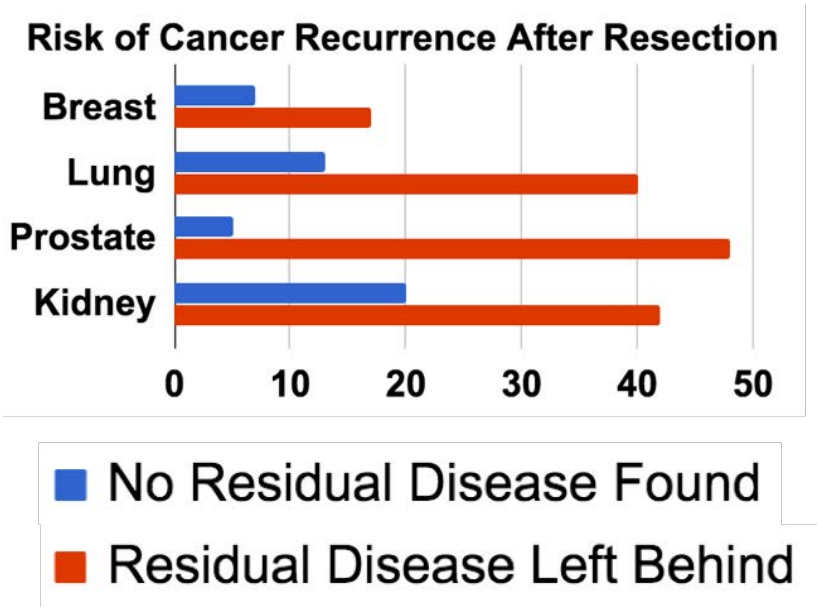


Figure 1. Effects of MRD on cancer recurrence rates.

all cancer types[2]–[4], necessitating substantial additional treatment[5], [6]. In addition to cost,

this increases risk of toxicity which can substantially reduce the quality of life[7], [8]. New technologies, such as targeted molecular imaging agents [9]–[11], miniaturized imaging platforms[12]–[15], and enhanced optical equipment, have been significant contributors to reducing the incidence of MRD. Nonetheless, MRD still remains a common occurrence[16], [17].

Once MRD has been assessed to have been left behind, it either necessitates a re-excision of the MRD or an acceptance of a higher risk of recurrence and death from cancer. Alternative strategies center around empirically removing an excessive amount of healthy tissue.

Despite the lack of a practical intraoperative imager, surgeons have tried to find other ways to identify microscopic residual disease or assess whether or not disease is left behind during surgery, but nearly all of the current methods require a significant amount of post-surgery lab analysis and the results often are too late to be practical. Current strategies for MRD detection rely on examining excised tissue labeled with highly specific cancer markers under a microscope, post-operatively. However, tissue processing, staining and microscopic evaluation of the excised tissue surface (which, by proxy, connotes the presence or absence of residual tumor on the corresponding area within the patient) restricts MRD evaluation to the post-operative pathology lab, taking days to weeks to return a result. This latency hinders the treatment process and compromises the success of the initial resection surgery in more than 25% of the cases, requiring additional surgical operations along with other forms of treatments such as irradiation, or chemotherapy. An efficient and practical intraoperative imaging platform is therefore an evident need for the successfulness of cancer treatment surgeries.

Recent medical advances in cancer therapy have also highlighted the importance of the ability to visualize MRD in lymph nodes that are often sitting several millimeters below the surface of the tumor bed. Being farther deep in, these lymph nodes are not able to be reached with current and commonly used excitation wavelengths for fluorescence imaging, which are mostly ultraviolet (UV) or visible, and as a result, being able to reach that deep into the tissue would require using molecular probes that can be excited with a longer wavelength of excitation light -which conventionally used fluorophores are unable to do. An ideal imager should therefore have the ability to perform deep tissue imaging using an excitation wavelength that can penetrate deep enough into the tissue to allow deep tissue and lymph node imaging as well.

1.2 Challenges

The miniaturization of large microscopes and fluorescent imagers require eliminating the necessary optics they rely on to resolve the image, i.e., color and wavelength filters and focusing lenses. Current intraoperative imaging apparatuses are too rigid, bulky and impractical due to their requirement for complex optics and filters. An ideal and practical intraoperative imager needs to be sensitive enough to be able to detect microscopic disease in real-time during resection surgery and provide a visualization of tumor foci in the tumor cavity for the clinicians, and additionally, have a small form factor and be compatible with clinical and surgical instruments often used during operations. In addition to that, it should have the ability to excite sites deep in the tissue to enable deep tissue and lymph node imaging for a more comprehensive therapy, which is possible using longer wavelengths such as NIR -while still maintaining an emission in the visible range to remain compatible with CMOS photodiodes.

Attempts at intraoperative image guidance continue to fail for MRD detection due to the bulk and rigidity of the required optics and filters. To help with identifying the extent and boundaries of the tumor and determining whether or not further resection of the tumor cavity and draining lymph nodes is necessary, current technologies leverage fluorescently tagged biomarkers

engineered to specifically attach to cancer cells, coupled with intraoperative imagers to visualize where the disease is present. Such an image resolution method requires optical filters which adds to the bulkiness of the overall imager and probe. Optical filters are required to reject excitation light which is often 4-6 orders of magnitude stronger than the weak fluorescent emission light, only shifted by 50 nm. Additionally, the issue of tissue and background autofluorescence remains a persistent problem and cannot be eliminated with optical filters. In addition to that, focusing optics are required to maintain spatial resolution and precisely guide the excitation and emission light through the optical system. These filters and lenses confer a size and rigidity impractical for a complete examination of the tight confines of modern minimally invasive tumor cavities or lymph node basins[18]. Furthermore, size-constrained to operate outside the tumor bed, far from the tissue, optical sensitivity is dramatically reduced.

Eliminating filters can be achieved by switching to a time-domain image acquisition where alternating pulses of excitation and acquisition windows are carried out while ensuring that the excitation light is no longer present during signal acquisition, therefore obviating the need for wavelength filters. However, conventional fluorophores have a very short phosphorescence lifetime (<10 ns) and are therefore not compatible with speeds CMOS imaging arrays operate at. Additionally, of the limited number of available fluorophores that can be excited in the NIR range and perform upconversion to emit in the visible range, none are efficient enough to be able to provide a strong enough signal to be detectable by CMOS sensors while complying with maximum permitted exposures and safety laser power limits.

1.3 Methods and Results Overview

This work presents a new strategy to move the cumbersome imaging requirements imposed by wavelength selective imaging to a time domain regime, inherently compatible with CMOS technology and large-scale pixel arrays. Enabled by uniquely efficient upconverting nanoparticles (UCNPs) [19]–[21] that have the ability to be excited in the NIR range, 980 nm and 1550nm, and emit in the visible range (with a mix of 550 nm and 660 nm light) by upconverting the input light using a 2-photon and 3-photon process respectively. In addition to being able to upconvert the excitation light, these nanoparticles have a very long emission lifetime (of several 100s of microseconds) and have been custom-engineered to be brighter than their conventional counterparts at lower powers -powers that meet the safety and maximum exposure limits of *in vivo* and clinical settings. A complete diagram of the upconversion process of the UCNPs is shown in **Figure 2**(a).

This thesis has culminated in the development of a standalone and fully-integrated sensor that has mitigated the limitations of large fluorescence microscopes and could be considered a reliable alternative for its cumbersome counterparts -a concept illustration of which is shown in **Figure 2**(b). We have eliminated wavelengths filters by using a time-domain image acquisition scheme and turning the excitation off during the window where emission is captured -a process illustrated in **Figure 2**(c). We have also integrated focusing optics directly on the chip by patterning micro-collimators fabricated using integrated metal layers and enhancing the angle selectivity of the sensor at no additional increase in form-factor.

The fully-integrated and optics-free presented here is an imaging platform freed from the need for optical filters and lenses, capable of being placed directly on the tissue surface. Spatial resolution has been enhanced using a combination of proximity and angle-selective gratings[22]–[24], integrated directly on chip. The resultant IC-only imager can be thinned to a negligible

thickness nearing only $25\mu m$, enabling a flexible scalable, molecular imaging skin, capable of being integrated on any surgical surface, such as medical or tissue biopsy probes.

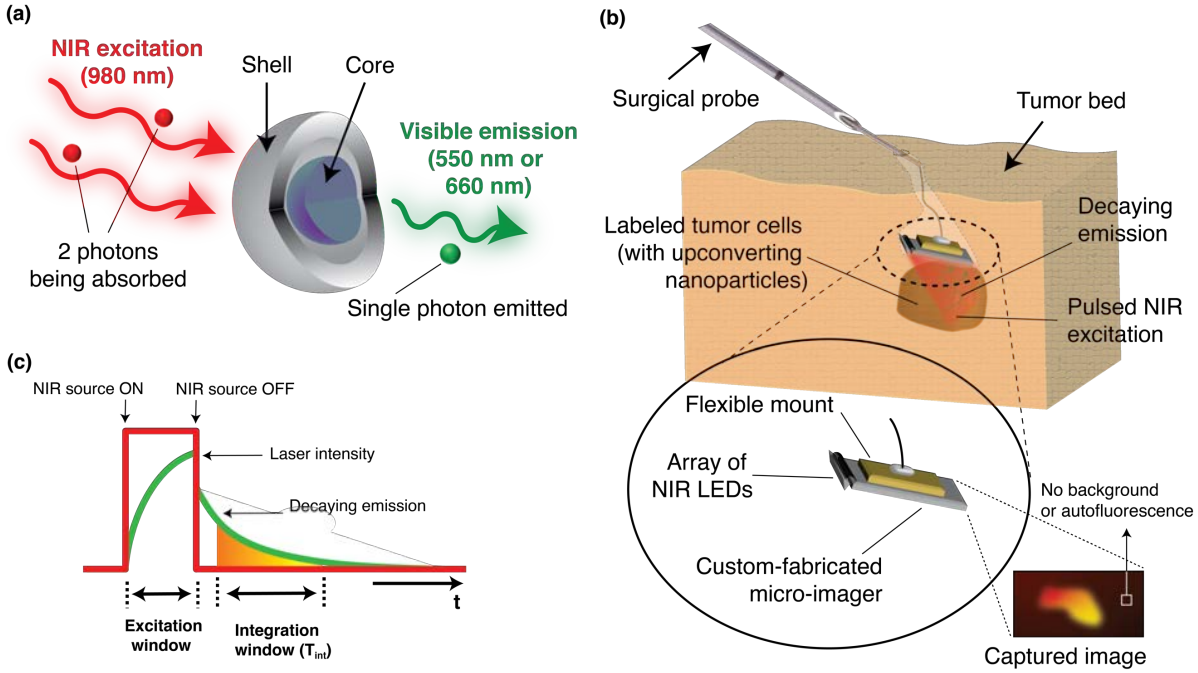


Figure 2. Concept overview of the proposed imaging platform: (a) illumination scheme of lanthanide-based upconverting nanoparticles. (b) proposed time-resolved optics-free intraoperative imaging platform for optically guided surgeries. (c) diagram of time-resolved image acquisition sequence.

2 Fluorescence Imaging

In this section, the theory and practical considerations of fluorescence imaging will be presented, and the two main methods of acquisition will be discussed.

While there are many flavors of fluorescence imaging currently being carried out, they are all based on the fact that biomarkers and fluorescent molecular tags, along with proper antibodies, can provide highly selective, very accurate and reliable binding to cancer cells and provide a distinct enough contrast between healthy and cancerous regions. Once attached to cancer cells, these molecular markers can be excited with a suitable excitation light source and the emission, once collected, captured and resolved, can provide us with the visualization of any microscopic cancer cell residuals.

When it comes to fluorescence imaging, there are two main approaches: frequency domain, and time domain image resolution. Each one differs in the way the excitation and emission light are respectively applied and extracted from the sea of signals generated around the sample.

2.1 Frequency domain resolution

2.1.1 Concept

Frequency domain resolution takes advantage of the fact that the emission coming from any fluorescent marker has in fact a different wavelength than the excitation light the particle reacts to and absorbs. Upon excitation with a fixed wavelength, a slightly longer Stokes-shifted wavelength (often only 30 to 50 nm longer) emission light is emitted by the targeted cells and captured by the imaging sensor, revealing any existing cancer cell.

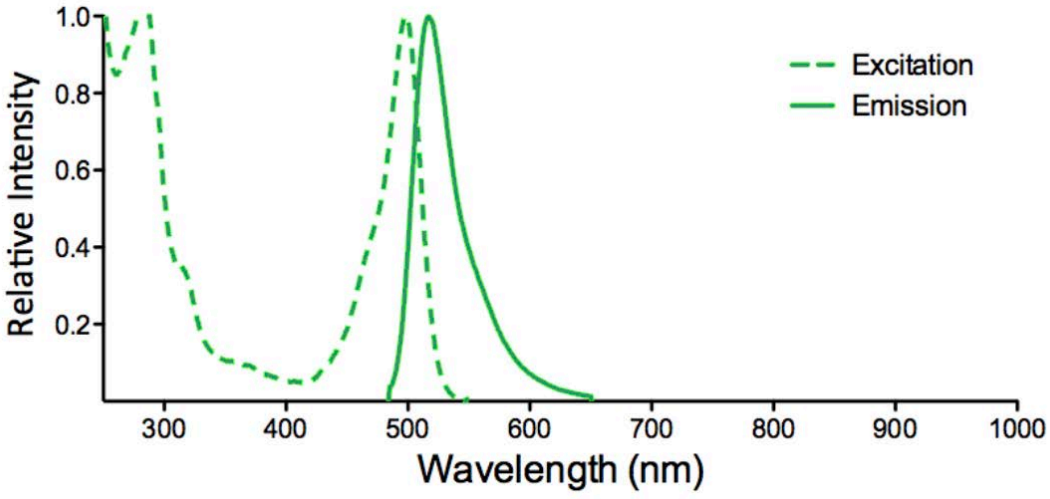


Figure 3. Excitation (absorption) and emission (fluorescence) spectra of FITC (courtesy of StressMarq Biosciences Inc.)

This shift in wavelength, while subtle and often only due to intrinsic losses in atomic transitions and photon generation, allows optical filters to separate and extract the emission signal from the excitation light. The excitation light is often many orders of magnitude (4 to 6 orders)

stronger than the emission, and this high selectivity requirement on optical filters is one of the greatest challenges of frequency domain fluorescence imaging.

Figure 3 shows an example of an organic fluorophore (Fluorescein or FITC) excitation and emission spectrum, illustrating how closely spaced the two spectra are, highlighting the extreme selectivity requirements imposed on optical filters.

2.1.2 Challenges

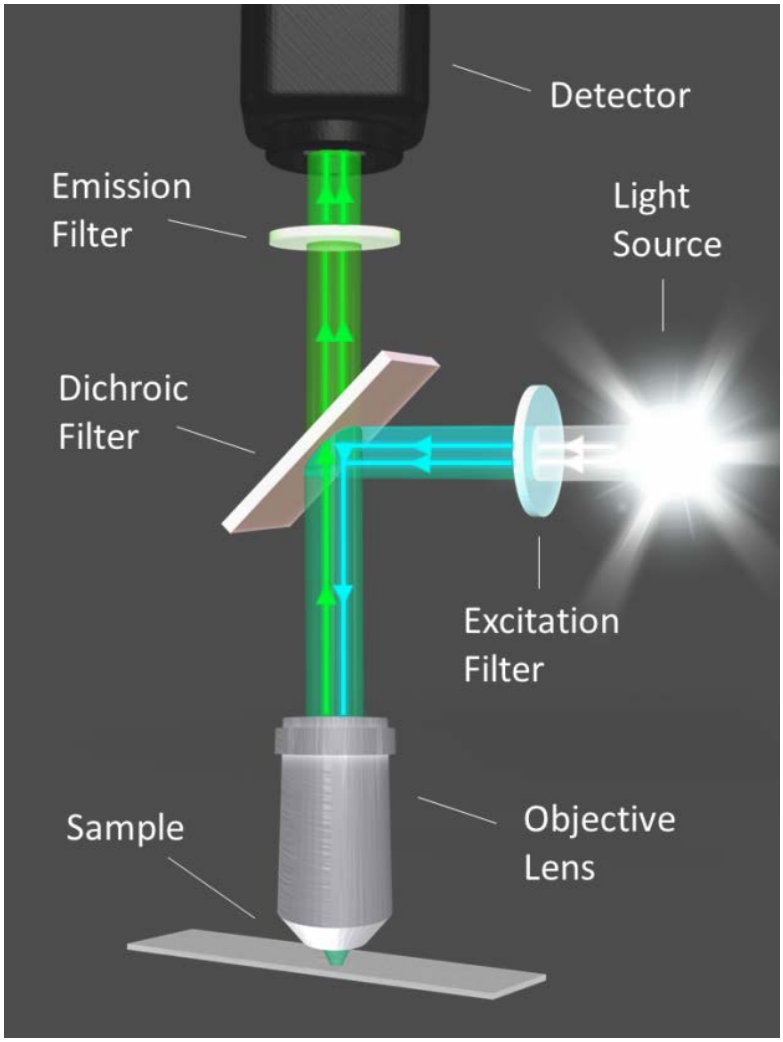


Figure 4. Diagram of a conventional frequency-domain fluorescence microscope imager (courtesy of Alluxa Inc.).

Frequency domain image resolution is heavily dependent on the accuracy and efficiency of the optical filter being used. The higher the rejection of the excitation light is, the better the signal-to-background ratio of the resulting image will be. Moreover, these optical filters will need to have an extremely narrow and sharp transition band to be able to separate two very closely spaced wavelengths, and will also have to be angle-insensitive to be able to let through the entire emission light emanating from the molecular markers, regardless of their angle of incidence -a feature that often cannot be guaranteed in commonly used filters. In addition to that, these filters will need to achieve at least 4 to 6 orders of magnitude of rejection to be able to extract the very

weak emission signal from the much stronger excitation light. The combination of all these stringent requirements and specifications for the optical filters makes frequency domain imagers bulky, cumbersome and unable to be miniaturized into smaller platforms with millimeter-scale form-factors, and thus, are not practical solutions for intraoperative imagers that are needed for use in today's minimally invasive surgical procedures and complex tumor cavities.

Figure 4 illustrates the optical elements and equipment needed to resolve the image using this imaging method and it can be seen that without any of these optics, resolving the image would not be feasible and these cumbersome elements precludes miniaturization beyond a centimeter scale range.

2.2 Time-domain resolution

Every phenomenon at the atomic level requires some finite amount of time to be fully carried out. Atomic transitions and photon generation and release are no exception. Such transitions require a given amount of time in which the electrons absorb the incoming excitation photon and release it back, generating the emission light in the process. As a consequence, if the excitation light is suddenly turned off, the emission light will still persist for a short amount of time afterwards, which enables time-domain image resolution to become possible.

2.2.1 Concept

Unlike frequency domain resolution, time domain image resolution does not rely on wavelength or frequency features of the excitation and the emission light. Instead, it is based on a time domain separation of the two lights, completely insensitive to frequency contents and therefore alleviating the need for any frequency-selective optical filter. In simpler terms, in order for the excitation light not to interfere with the capturing of the emission signal, the excitation light is turned off during the image resolution process, no longer necessitating an optical filter to reject and block the interfering excitation light. Once the excitation light is turned off and the emission light starts to decay, the remaining emission signal is then collected and captured by the sensor.

2.2.2 Challenges

The main challenge with a time-domain image resolution technique is the fact that most of the conventional fluorophores and molecular markers available have a very short emission lifetime (only a few nanoseconds), which imposes an extremely heavy limit on the speed of the image sensors being used in conjunction with those particles[25]. With a lifetime of only a few nanoseconds, the sensor and required circuitry would need to operate at a significantly higher speed than conventional silicon sensors are capable of. In addition, care should be taken that the excitation light does not interfere with the normal course of operation of the silicon imager and compromise the signal-to-background ratio. Since CMOS sensors are virtually only comprised of silicon structures, they will be affected by any excitation light that has high enough energy to create carriers (electron/hole pairs) in the active region or substrate and introduce unwanted interference on the pixels. These carriers will have a recombination lifetime of their own when created, and one of the main challenges of time-domain acquisition is to mitigate the effect of this background generated by the excitation light.

In addition, the lifetime of tissue autofluorescence is often virtually indistinguishable from the desired emission[26], significantly limiting the achievable signal-to-background ratio (SBR)

for lifetime imaging. Unlike excitation interference, tissue and cellular autofluorescence –which is more pronounced with shorter wavelengths such as UV– cannot be completely blocked or removed by filtering. Longer emitting biomarkers with lifetimes longer than a few microseconds and that are excited at longer wavelengths, in the NIR-I (700 -1000 nm) or NIR-II (1000 - 2000 nm) range, have the ability to address both of these issues[27].

3 Upconverting Nanoparticles

Upconverting nanoparticles, unlike conventional fluorescent markers, are capable of sustaining an emission in the visible range, while being excited in the near-infrared (NIR) range[19]–[21], [28]. We have investigated three particles in this work, each with a different composition, in order to characterize their relative intensity and emission lifetime. These three nanoparticles have NaYbErF4 as their core, but each has a different composition ratio of Yb/Er; 20/80, 60/40 and 80/20. Yb is the main photon- absorbent at 980 nm, while Er is more efficient at absorbing 1550 nm light, where 980 nm and 1550 nm are the two excitation wavelengths for a 2- and 3-photon process respectively. All three UCNPs have 8 nm sized core, with 4 nm sized shells of $NaYF_4$ with 20% Gd, diluted to a concentration of 0.68 μM in Hexane. A negative control and reference is also measured for each experiment using a Hexane-only solution.

Specifically engineered and with the correct linker molecule and antibody, they can bind to tumor cells allowing greatly enhanced selectivity compared to non-specific biomarkers. Furthermore, these UCNPs do not suffer from photobleaching – major drawback of conventional fluorophores. As a result, these nanoparticles retain their optical characteristics over time and do not degrade with repeated or long imaging periods. In addition to being immune from photobleaching, thanks to their long and low energy excitation wavelengths (980 nm and 1550 nm), these particles can create a regime where no autofluorescence or tissue background is present, and a higher signal-to-background ratio (SBR) can be achieved.

To quantify and compare the three different compositions of UCNPs, we have measured their power efficiency, emission lifetime and intensity using our custom-made CMOS imager, at both absorption wavelengths (980 and 1550 nm). A vial containing 400 μL of the 0.68 μM UCNP solution is placed directly above the imager array, and the solution is then illuminated (excited) with the corresponding time-gated collimated laser beam. **Figure 5** depicts the experiment setup for these measurements.

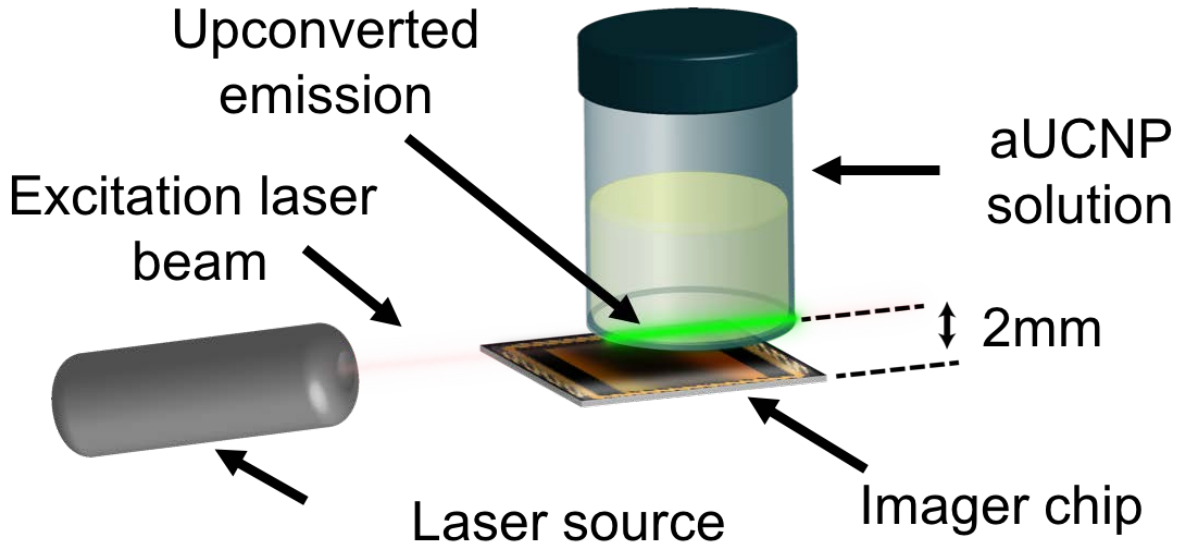
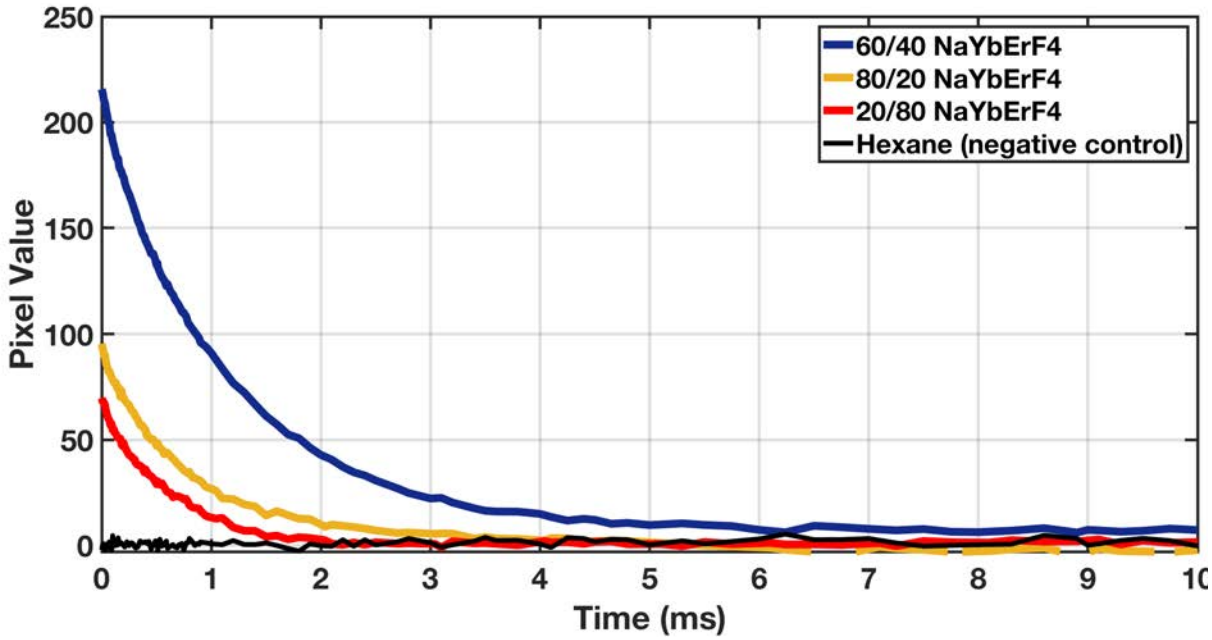


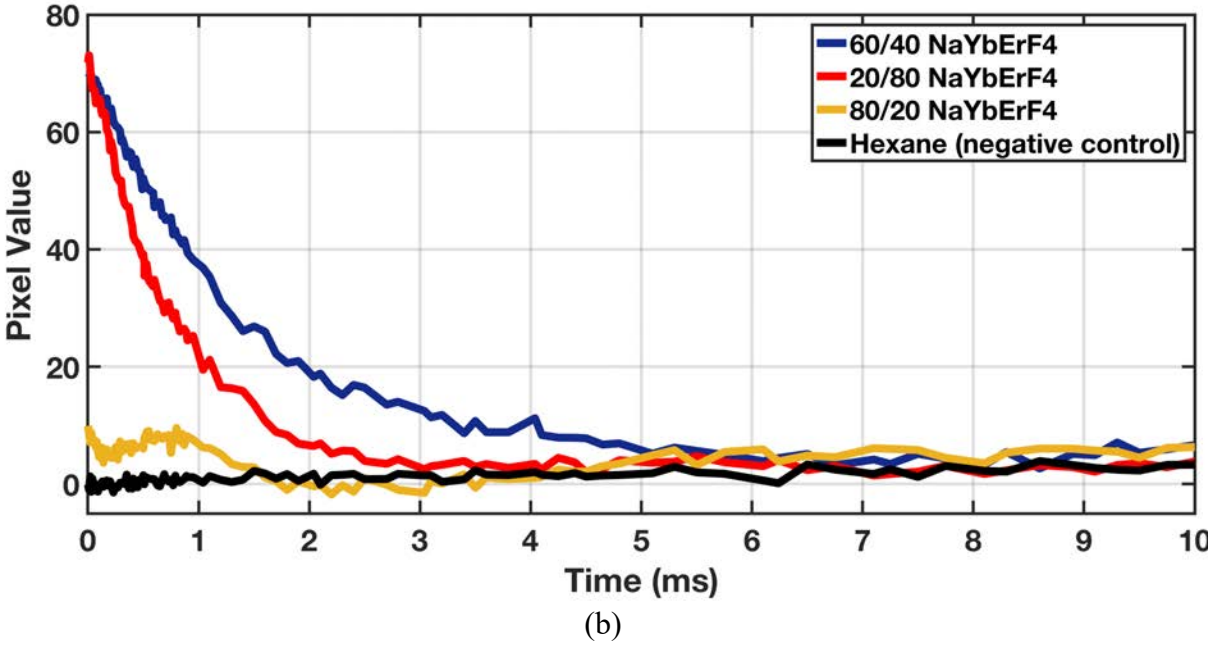
Figure 5. Experiment setup for UCNP characterization.

The three main factors contributing to the compound efficiency of the UCNPs for clinical application are power conversion efficiency, emission lifetime and the duration of excitation required[29].

3.1 Emission decay

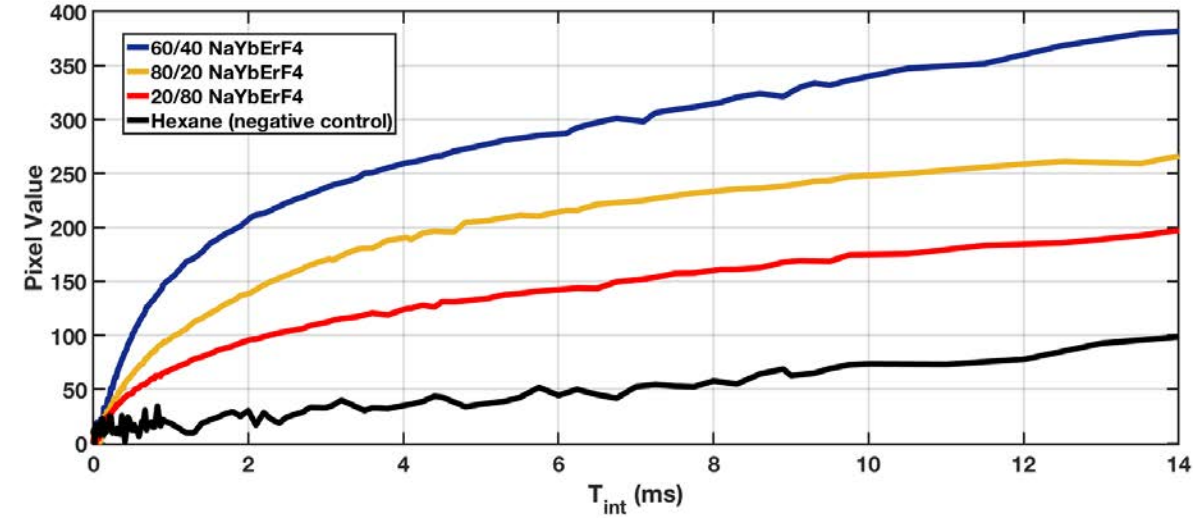


(a)

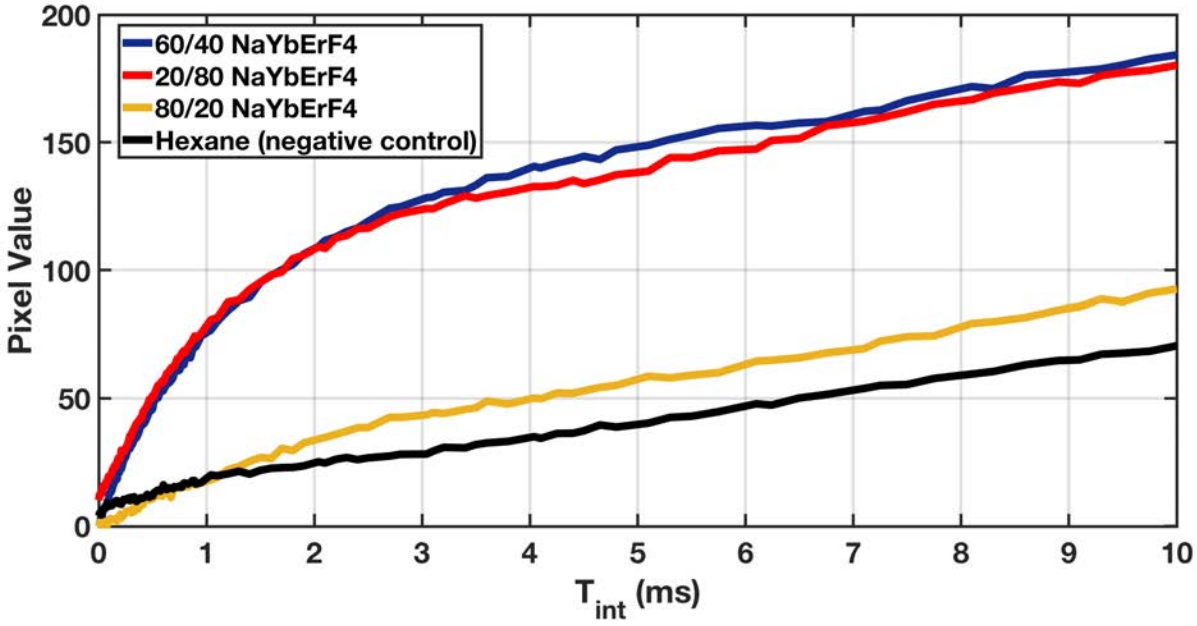


(b)

Figure 6. Measured emission intensity decay of UCNPs (dark current subtracted and T_{exc} = 5 ms):
(a) at 8 W/cm^2 of 980 nm excitation. (b) at 60 W/cm^2 of 1550 nm excitation.



(a)



(b)

Figure 7. Measured integrated emission intensity of UCNPs (T_{exc} = 5 ms): (a) at 8 W/cm^2 of 980 nm excitation. (b) at 60 W/cm^2 of 1550 nm excitation.

To obtain the emission decay lifetime, we can reproduce a representation of the decay profile using a moving integrating window, easily realizable with the custom-made imager. As derived in [29], the measured photocurrent will be as follows:

$$I_A(t) = I_{A_0} e^{\left(-\frac{t}{\tau}\right)} + I_D \qquad (1)$$

where τ and I_D are respectively the emission decay lifetime and dark current level in the pixels. **Figure 6**(a) and **Figure 6**(b) show the measured emission decays for the three UCNPs and the negative control, at 980 nm and 1550 nm excitation respectively. The dark current level is subtracted from the waveform for visual purposes. The data points have been obtained using a

moving but fixed-length integration window of 1 ms. The emission profiles have been measured under a time-gated excitation light generated by a fiber-coupled laser, with the corresponding wavelength. The instantaneous power for the 980 nm and 1550 nm laser used was 8 W/cm^2 and 60 W/cm^2 respectively.

The emission decay profiles obtained at 1550nm excitation visibly show that compared to the 980nm-excited 2-photon process, the 3-photon process results in an approximately 22 times less brighter emission signal. Keeping in mind that silicon's bandgap energy corresponds to a wavelength of 1103 nm[30], despite the apparently poorer upconversion of a 3-photon process, the overall efficiency of the imaging system can be higher, owing to the fact that a 1550 nm excitation source will introduce negligible interference on the Si-based CMOS circuit compared to its 980 nm counterpart. Using higher excitation powers at 1550 nm can help compensate for the smaller upconversion rate, while making sure that the system is still fully compliant with maximum exposure limits and safety standards.

Another quite useful way of comparing the intensity of the three UCNPs is to look at the integrated emission intensity, that is the total measured signal level at different integration window lengths. **Figure 7**(a) and **Figure 7**(b) illustrate these results. Derived in [29], the obtained results can be expressed as

$$I_B(t) = I_{B_0}\left[1 - e^{\left(-\frac{t}{\tau}\right)}\right] + i_d t \quad (2)$$

where I_{B_0}, τ and i_d are respectively a constant, the emission decay lifetime and the dark current intensity. The second term of this equation ($i_d t$) represents the dark current which can also be observed in all the plots in **Figure 7**. As shown in **Figure 7**, the Hexane negative control has the lowest pixel values, and given the absence of any UCNPs, their measured integrated signal is a sloped line with no exponentially decaying component.

3.2 Upconversion efficiency

To characterize the upconversion efficiency of the UCNPs, we have measured the upconverted emission at different power intensities. The measured emission intensities are plotted in **Figure 8**. The measurements have been conducted at both excitation wavelengths and compared with respect to the Hexane negative control.

The upconverted emission is measured to be quite linear with the input excitation power, for both the 2- and 3-photon process. At the relatively lower powers where the measurements are conducted at, the output flux of emission photon is measured to be, to first order, linearly proportional to the input flux. Higher power levels may lead to secondary order non-linearities which are not applicable and relevant to our case.

Comparing the data for the two processes in **Figure 8**, it can be seen that the 980 nm light produces a small amount of interference, reflected in the Hexane negative control plot in **Figure 8**(a). This interference is generated through the scattering of the excitation laser beam onto the Si-based imager, resulting in an undesired increase in the background level. However, such an interfering effect is almost negligible and not visible in the measurements, as seen in **Figure 8**(b).

3.3 Excitation duration

Another contributing element to the signal intensity is the duration of excitation, T_{exc}, which represents how long the nanoparticles are illuminated with the excitation light source before

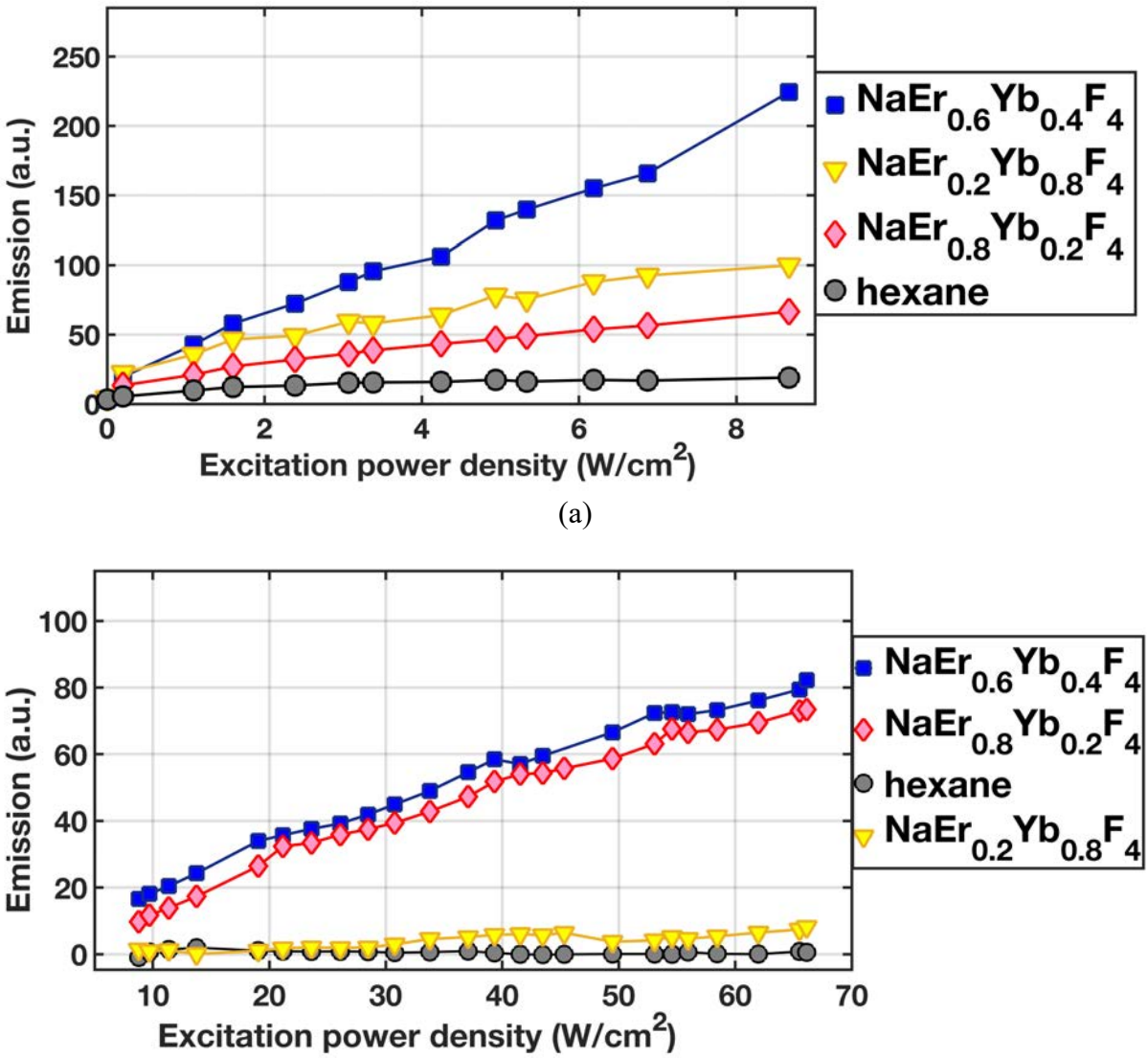
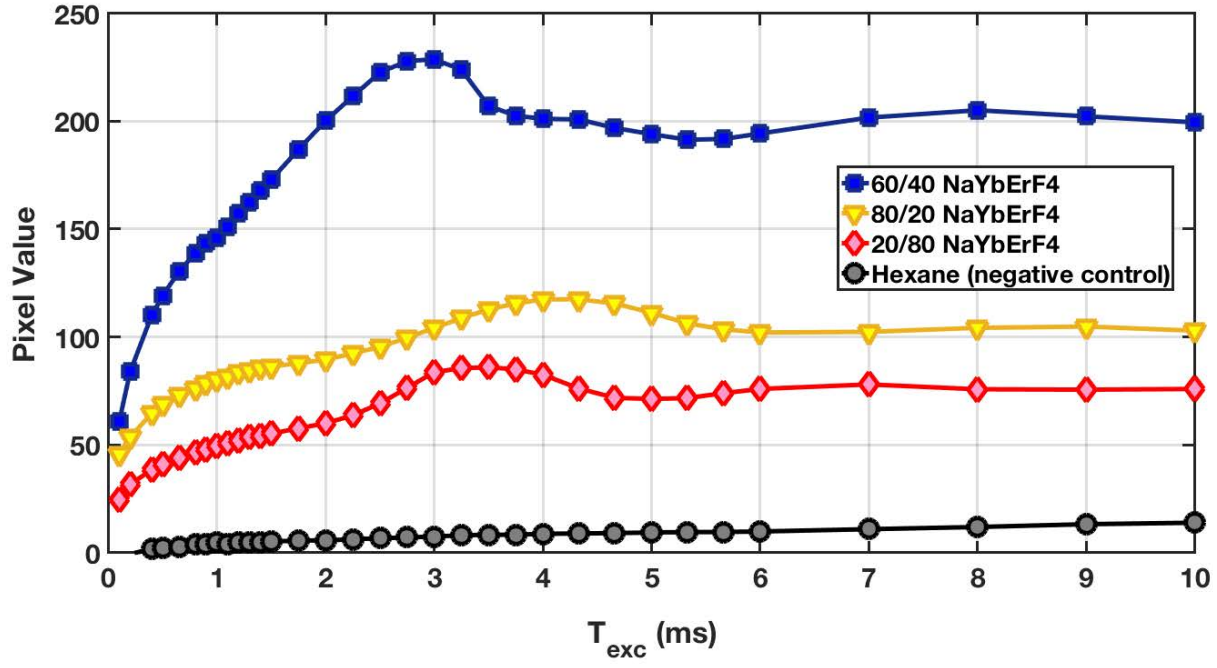
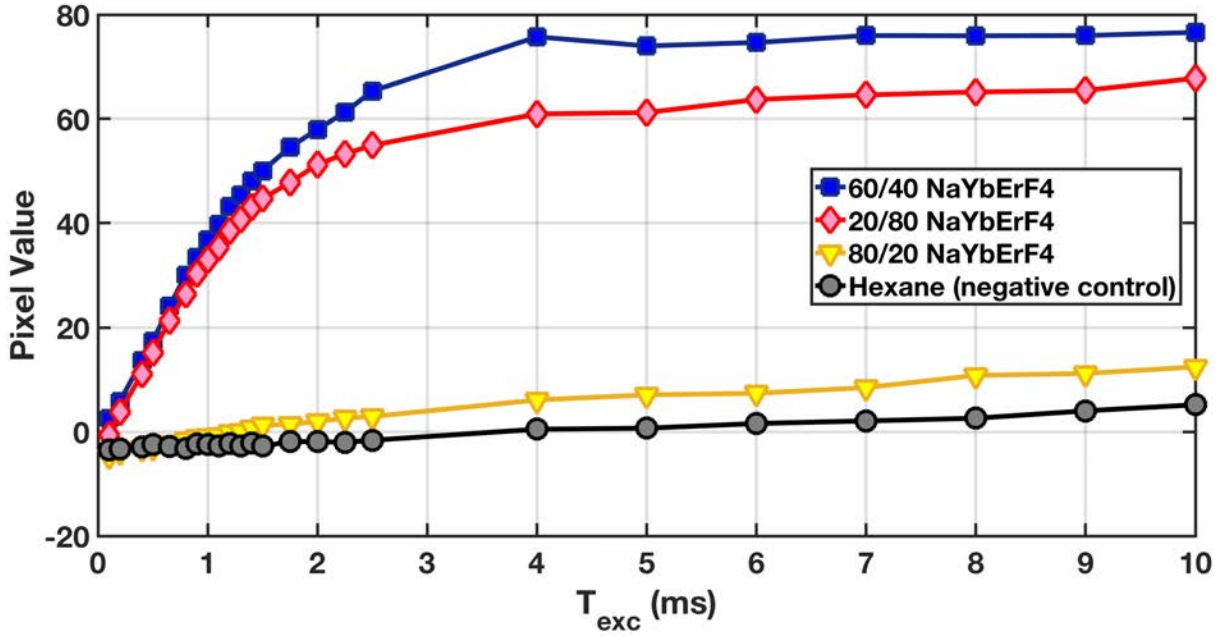


Figure 8. Measured emission intensity of UCNPs at different power levels ($T_{int} = 1\ ms, T_{exc} = 5\ ms$): (a) 2-photon process with 980 nm excitation. (b) 3-photon process with 1550 nm excitation.

the start of the time-gated imaging sequence. The UCNPs were excited for various durations of time (T_{exc}) and the measured emission intensities are brought in **Figure 9**(a) and (b) for 980 nm and 1550 nm excitation respectively. It can be seen from the plots in **Figure 9** that there is a noticeable increase in the emission in all three UCNPs, until T_{exc} reaches a point after which no significant increase is observed. For our work, we have used a 5 ms illumination time for our experiments (T_{exc}= 5 ms). It should be noted that being externally driven with a synchronized pulse stream, the excitation laser power is susceptible to any transient non-idealities in the pulse waveform, i.e., ringing effect, as seen in **Figure 9**(a).



(a)



(b)

Figure 9. Measured emission intensity of UCNPs at different power levels:
(a) 2-photon process with 980 nm excitation. (b) 3-photon process with 1550 nm excitation.

4 Imaging ASIC With Dual-Photodiode Pixel Architecture

Using a long excitation wavelength (e.g., NIR) has the advantage of virtually eliminating all tissue and background autofluorescence, and as a result, NIR imaging has always been a very appealing field in fluorescence microscopy. However, the impact and interference an NIR excitation light generates on a CMOS-based imaging sensor needs to be considered as well. Although no background illumination light or autofluorescence remains, when pulsed, NIR excitation light still results in large amounts of interference in CMOS-based photodiodes. This occurs because the NIR illumination light, absent a filter, will penetrate into the silicon bulk, beneath the photodiodes, generating substrate carriers with diffusion times on the order of hundreds of microseconds[31]. This introduces a background current in the photodiode that needs to be measured and subtracted from the desired signal. We here present a time-gated lens-less and filter-less chip-scale solution that mitigates the interference and background challenge by incorporating a dual photodiode pixel architecture and localized background level and interference adjustment method.

In this section, challenges of imaging using NIR wavelength excitations (without optics) will be discussed, and our developed novel dual-photodiode architecture will be described, and the experimental measurement and verification of its performance will be presented afterwards.

4.1 Challenges of NIR imaging

Time-gated illumination schemes obviate, in principle, the need for optical emission filters since the excitation light is turned off during image acquisition. However, the main challenge arises when light with a wavelength that travels deeper into silicon (i.e., NIR light such as 980 nm) penetrates into the bulk silicon generating a significant number of electron-hole pairs which enter the photodiode and function as interfering carriers. As seen in **Figure *10***(a), these carriers are generated throughout the substrate, but the low doping of the bulk enables a slow recombination, allowing them time to travel towards the depletion regions at the surface where they recombine, thus creating a background interference on the photodiodes. Given the angular and spatial uniformity of the carriers' paths inside the substrate, **Figure *10***(a) also illustrates how these interfering carriers will even reach photodiodes covered by metal layers -albeit to a different extent. As a result of the low number of dopants present in the substrate, this undesirable background will have several hundreds of microseconds of recombination lifetime, posing a great challenge for time-gated acquisition.

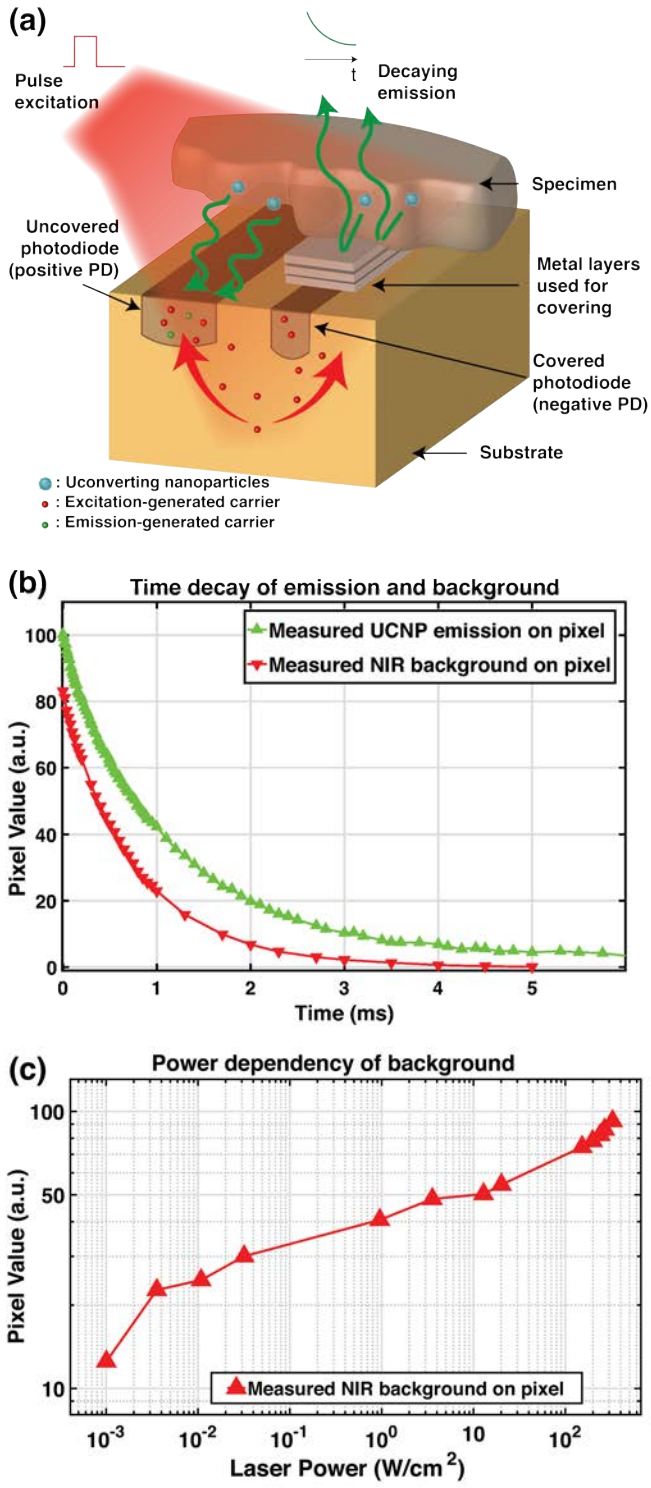


Figure 10. Effect of pulsed NIR excitation on covered and uncovered photodiodes: (a) positive (uncovered) and negative (covered) photodiode and the locations and paths of the two different kinds of charges generated within the pixel. (b) time domain characteristics of UCNP emission decay and background generated by the pulsed excitation light on pixels (5-ms pulsed 980nm laser at 180 W/cm^2). (c) measured background level on the PPD as a function of the pulsed (time-gated) 980nm excitation light (pulse duration = 5 ms).

While NIR light will stimulate a photodiode whether covered or not, shorter wavelengths such as UV or the visible emission of UCNPs that do not penetrate deep into the silicon, are more easily blocked by metal layers, and mainly create optically induced charges only in uncovered diodes, as seen in **Figure *10***(a). Thus, we can use this distinction to measure the local background generated and subsequently remove it from the uncovered photodiode. The measured intensities and decay times of both UCNP emission (with a total particle size of 26 nm, comprising a 16 nm $NaEr_{0.8}Yb_{0.2}F_4$ core and a 5 nm $NaY_{0.8}Gd_{0.2}F_4$ shell[21] and concentration of 0.68 μM) and interfering background on a pixel are shown in Fig. 8(b). The UCNP emission has an effective decay lifetime of 950 μs, very similar to the 900 μs average recombination lifetime of interfering carriers, and their similarities in intensity and decay make it challenging to distinguish the signal from the background. Importantly, the intensity of the background does not scale down linearly with the excitation power, as shown in **Figure *10***(c) demonstrating that even reducing illumination power dramatically (such as with a thin-film interference filter) does not eliminate the background notably. The converse is also true, that by increasing the optical power, we do not substantially increase our background, but do increase our signal proportionally.

This NIR-induced background will interfere the desired signal and even virtually mask the desired emission if the background is strong enough. As a result, we have developed a novel pixel architecture to mitigate this challenge and help recover the underlying emission data from the compromised photodiode signal.

4.2 Dual-photodiode pixel design

To mitigate the issue of background induced by the pulsed NIR excitation light in the CMOS photodiodes, in our developed imaging system, we have included a novel background correction method that leverages a secondary and fully covered photodiode labeled “negative photodiode” (NPD), shown in **Figure *10***(a), to extract a local and per-pixel measurement of the background and adjust the primary and uncovered “positive photodiode” (PPD) background accordingly and reduce the residual background to near noise level.

The addition of a second photodiode will result in overall signal loss due to a reduced fill factor. In the case of equal pixel area allocation for the two photodiodes, the system would see its dynamic range lowered by 6dB. As a result, the size of the secondary diode should only be as large as needed, and instead additional ratio adjustment is preferred to translate background levels between the two photodiodes. **Figure 11**(a) illustrates the implementation of the two photodiodes, demonstrating their dissimilarity in both size and shape. The PPD is implemented using 4 identical 19 μm-by-19 μm photodiodes in parallel, covered by integrated angle selective gratings[23] to block oblique light from reaching the pixel to enhance spatial resolution as described in [23]. The NPD is formed from a cross-shaped active region spanning the space between the 4 PPDs, overall providing a centroid formation for a more uniform background measurement.

As seen in **Figure 11**(a), the two diodes are not identically sized, and therefore ratio adjustment is needed, where ratio is defined as

$$f(B_N) = \frac{B_P}{B_N} \qquad (3)$$

with B_P and B_N being the background levels measured on the PPD and NPD respectively. The characteristics of $f(B_N)$ were measured by extracting the ratio at various background levels for each individual pixel, and the aggregate of all ratios extracted for the entire array is shown in Fig. 9(b). It can be seen from **Figure 11**(b) that there is a non-linear dependency in the ratio and needs therefore to be accounted for when using the NPD to adjust the background of the PPD. To this

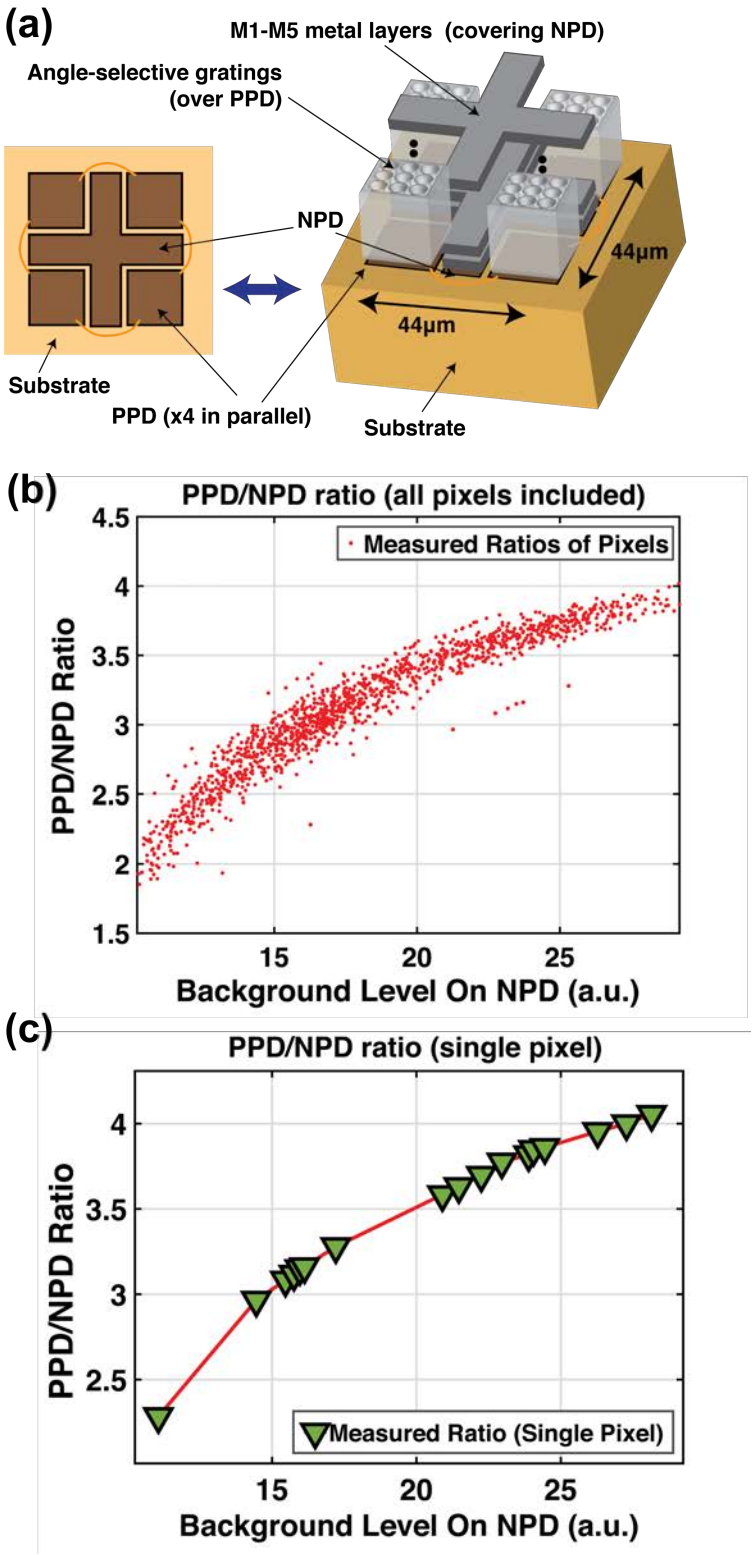


Figure 11. Non-linear dependency of positive-to-negative photodiode ratio: (a) 3D diagram of PPD (covered by angle-selective gratings) and NPD (covered with M1-M5 metal layers) in the pixel. (b) aggregate scatter plot of the ratios for all the pixel at various samples of NPD background levels. (b) fitted curve of the photodiode ratio as a function of measured NPD background for one sample pixel.

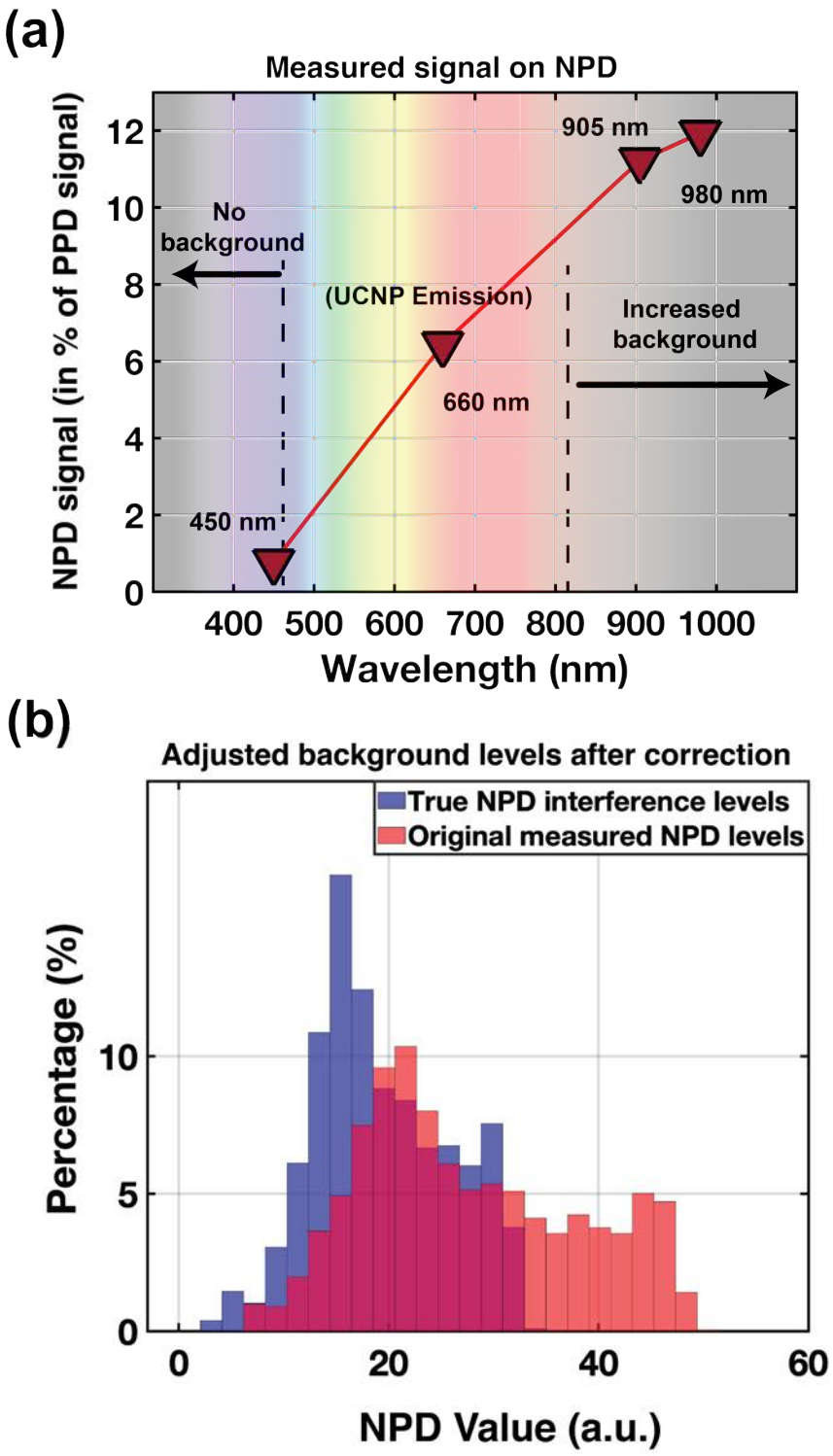


Figure 12. Overview of NPD correction and extraction of true background. (a) background generated on NPD at different wavelengths (namely 660 nm, which is of interest) as a percentage of the background on PPD. (b) histogram of original NPD values and true NPD background levels extracted using pixel-level non-linear curve fitting (the data for these two histograms were captured from an image of a 0.68 µM UCNP dispersion chamber, placed directly on the chip, and illuminated with 5 ms pulses of 980 nm excitation light at 45 W/cm^2).

end, the data for each individual pixel PD-pair is used to perform a one-time ratio curve fitting to extract the profile of this ratio. A curve fitting example for a sample pixel is shown in **Figure 11**(c). This profile extraction is a one-time operation and does not need to be repeated with every operation cycle, and therefore does not create a significant overhead to the image acquisition process.

Despite expecting the covering of the secondary photodiode to act as an optical shield for emission signals, additional investigation revealed that the 5 metal layers did not exhibit a perfect rejection of the emission light. Those metal layers used to cover the NPD from the emission light, almost completely shield the NPD from short wavelengths such as UV, and thus the NPD level measured will only include the background generated by the excitation light, easily removed with ratio adjustment. However, the performance of the blocking metal layers starts to degrade with longer wavelengths, and at 660 nm -the UCNP peak emission wavelength- 6.5% of the emission signal captured by the PPD is also present in the measured signal on the NPD, as seen in Fig. 10(a). Failing to mitigate this "leakage" will result in not only an emission signal loss after adjusting the PPD background, but due to the non-linear dependency of the ratios of the two photodiodes, the background adjustment will also be corrupted.

To account for the emission component measured on the NPD, we parametrize the measured signal on both photodiodes using the following equations:

$$\begin{cases} \mathrm{PPD} = \mathrm{E_P} + \mathrm{B_P} \\ \mathrm{NPD} = \mathrm{E_N} + \mathrm{B_N} \end{cases} \tag{4}$$

where E_P and E_N are the emission components of the measured values on the PPD and NPD respectively, and B_P and B_N are the excitation background generated on the PPD and NPD. Knowing the proportional (linear) relationship between E_P and E_N (from Fig. 10(a)) and the ratio $f(B_N)$, as defined in (3), we can further simplify (4) into

$$\xrightarrow{\mathrm{B_P} = \overbrace{B_N}^{unknown} \times \mathrm{f(B_N)}} \begin{cases} \mathrm{PPD} = \mathrm{E_P} + \mathrm{B_N} \times \overbrace{\mathrm{f(B_N)}}^{PD\ ratio} \\ \mathrm{NPD} = \mathrm{kE_P} + \mathrm{B_N} \end{cases} \tag{5}$$

where k is the proportional coefficient extracted from **Figure 12**(a). While k is wavelength dependent (see **Figure 12**(a) for dependency), we may consider a constant value thanks to the fact that majority of the emission power of the UCNPs is concentrated around 660 nm. As a result, using the measured values of PPD and NPD, we can then solve (5) for B_N and extract the true and corrected level of background. Removing $\mathrm{B_P}$ from the PPD signal then becomes trivial using (4) and $f(B_N)$. **Figure 12**(b) shows the distribution of a sample set of NPD measurements and the corrected and extracted background levels and highlights the importance of this correction.

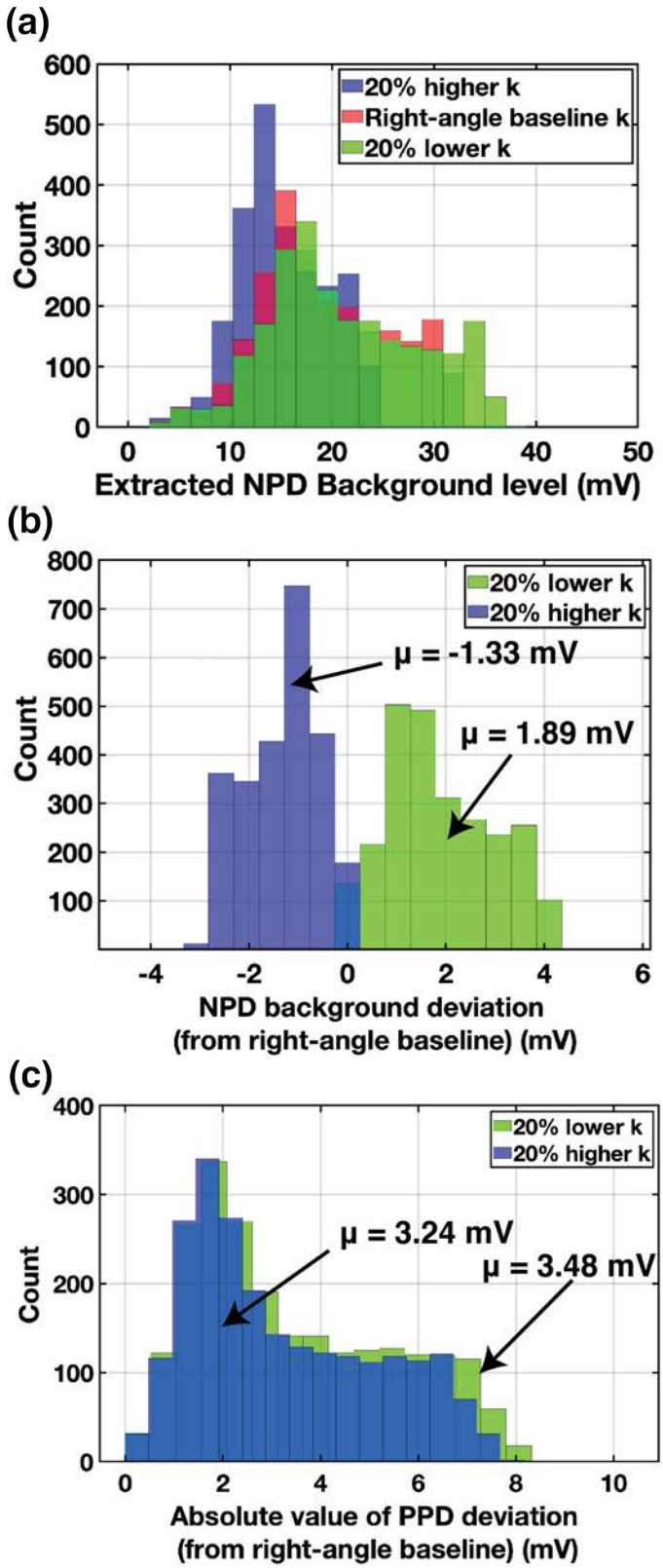


Figure 13. Effect of angle dependency, and k variation on NPD and PPD correction reliability. (a) histogram of sample set of extracted and adjusted NPD background values at baseline k, 20% higher k and 20% lower k value. (b) histogram of the deviation of adjusted NPD background with 20% higher and lower k value from their baseline values. (c) absolute value of effective deviation of PPD levels from their baseline values with a 20% higher and lower k value.

Precision of this technique relies on the accuracy of the proportional coefficient k that is used. While the values shown in **Figure 12**(a) are measured at normally incident light, it is evident that the emission light coming from an arbitrary specimen on the imager will not be confined to a normal incidence. To quantify the limits of this correction, the PPD background levels were extracted using a 20% larger and smaller k, to account for its angle-dependency with a reasonable range of change in its value. **Figure 13**(a) shows a histogram of the extracted background values in the 3 cases of a 20% higher, baseline (normal incidence) and 20% lower k value. The individual

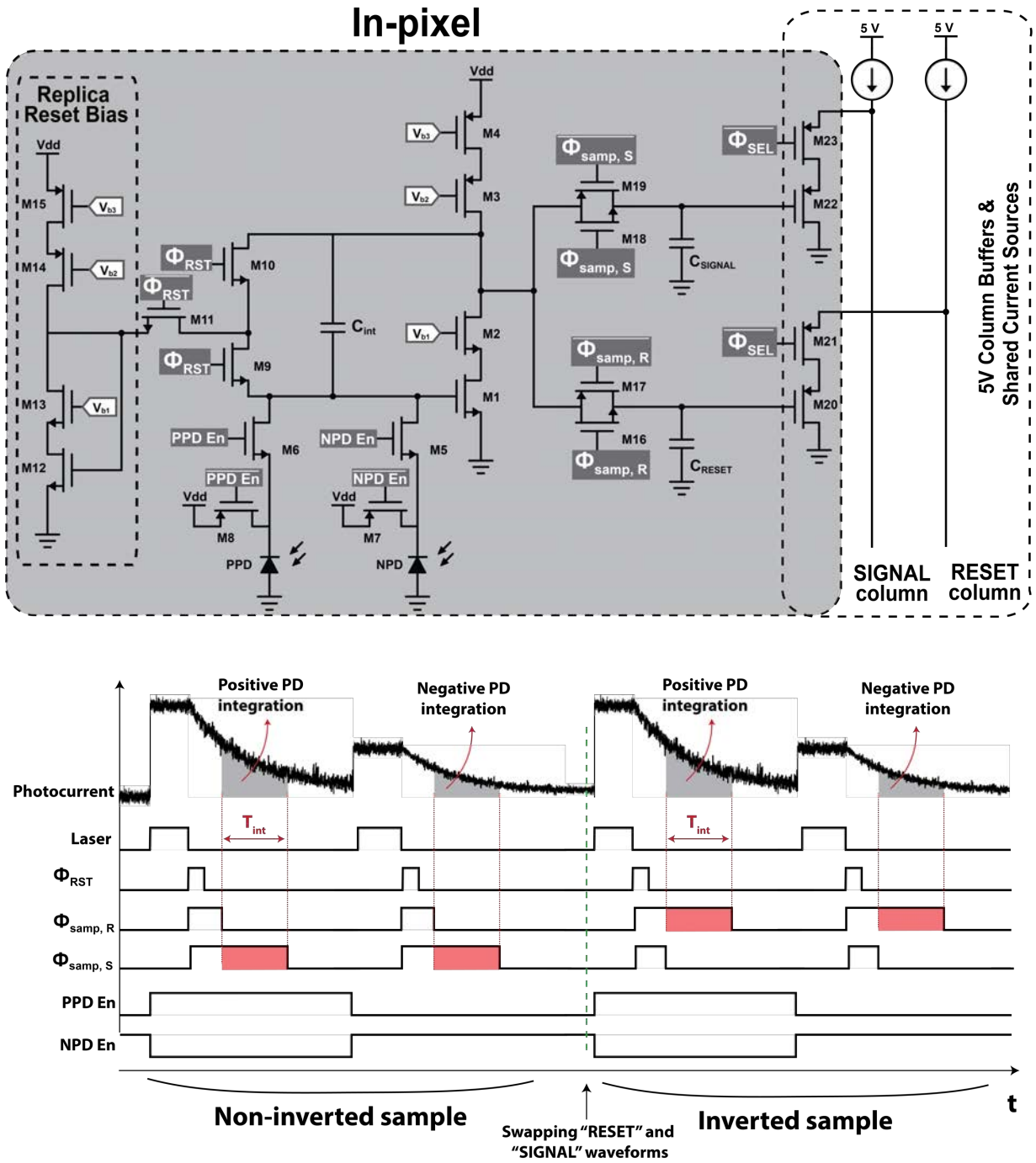


Figure 14. Schematic of the pixel circuit and control signal waveforms.

deviations of the NPD backgrounds extracted from the baseline (normal incidence) case is shown in **Figure 13**(b), demonstrating a mean NPD background deviation of -1.33 mV for a 20% higher k and 1.89 mV for a 20% lower one. The PPD deviations from their baseline values is shown in **Figure 13**(c), demonstrating that the overall correction precision of this method is near 3.5 mV.

The circuit schematic of the pixel front end is presented in **Figure 14**. There are several challenges to achieve a robust and linear pixel design that need to be carefully addressed, such as a sufficiently large dynamic range, a linear optical response, signal-dependent leakage of the switches, flicker noise and fill factor loss. The pixel architecture is a capacitive transimpedance amplifier (CTIA) which provides an acceptable dynamic range and a linear response, provided its loop gain is sufficiently high.

The integration capacitor Cint placed in feedback is a custom metal-oxide-metal (MOM) capacitor with a capacitance of 11 fF. This highly linear capacitor and a high gain front-end amplifier ensures a linear response to optical stimulation over the entire dynamic range. The two photodiode paths also include switches (M7-M8) to connect the photodiodes to the analog supply whenever not in use, to ensure leakage across the switches M5 and M6 remains signal-independent. To maintain the leakage across the integrating capacitor's reset switch (M9-M10) signal-independent as well, a replica biasing circuitry is connected when the switch is opened (when ϕ_{RST} is low), to keep a constant voltage applied to the switch[23]. This allows the leakage of the switches to appear as a constant offset at the output, which can be easily measured and cancelled out during acquisition.

Each photodiode's signal is measured individually in a time-domain duplexed manner, and the data is acquired using two rounds of correlated double sampling (CDS) schemes. **Figure 14** shows the timing of the control waveforms, where during the non-inverted acquisition for either

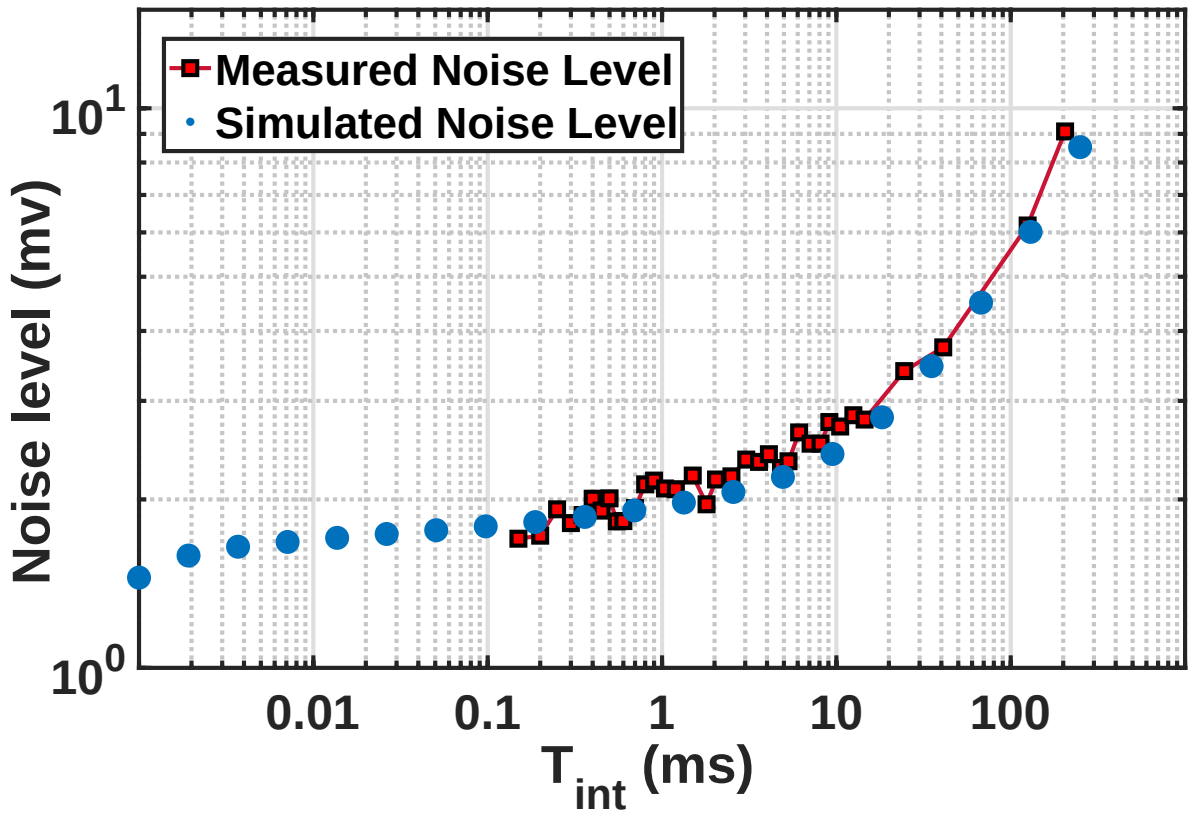


Figure 15. Pixel noise measurement and simulation results.

of the photodiode, the laser light is initially pulsed (for 5 ms) and upon turning off, the acquisition begins with ϕ_{RST} going high (for about 100 µs) and resetting the charge on the integration capacitor. The first CDS scheme, implemented to mitigate the effects of flicker noise, is carried out by capturing two samples, "RESET" as the baseline and "SIGNAL" as the secondary CDS sample, each of which is first stored on a register capacitor and transferred to the output through dedicated column buffers. The integration time is determined by the spacing of the "RESET" and "SIGNAL" acquisition timepoint, as shown in **Figure 14**.

The second CDS scheme is included to remove systematic offsets in the "RESET" and "SIGNAL" buffering paths and is implemented by acquiring a second sample for each photodiode, with swapped "RESET" and "SIGNAL" register capacitors, resulting in an inverted sample (shown in **Figure 14**). This offset will appear in the two samples as a common-mode component which can be removed, and the underlying signal be retrieved via simple subtraction. Thanks to its fast readout time of 1.2 ms and repeatability, a single acquisition can be repeated as many times as needed and when averaged, increases the SNR by a factor proportional to $\sqrt{N}$, where N is the number of repetitions.

The noise performance of the pixel was simulated at various integration times, with all sources of noise sources accounted for, including but not limited to shot noise, thermal noise, and flicker noise, which are the major contributors of noise to the system. The shot noise in particular is a signal dependent noise which increases whenever emission is present and therefore a larger photocurrent is generated, and is therefore the dominant source of noise at longer integration times (above 10 ms). For simplicity and without loss of generality, all noise measurements and simulations have been performed in the absence of emission and excitation signal. The noise of the pixel was measured, and **Figure 15** shows the results of the simulated and measured noise values. For a typical integration window duration of 1 ms, the measured noise voltage was 2.2 mV rms, and is dominated primarily by the thermal noise of the front-end amplifier.

4.3 Time-domain averaging and impact on measurement noise

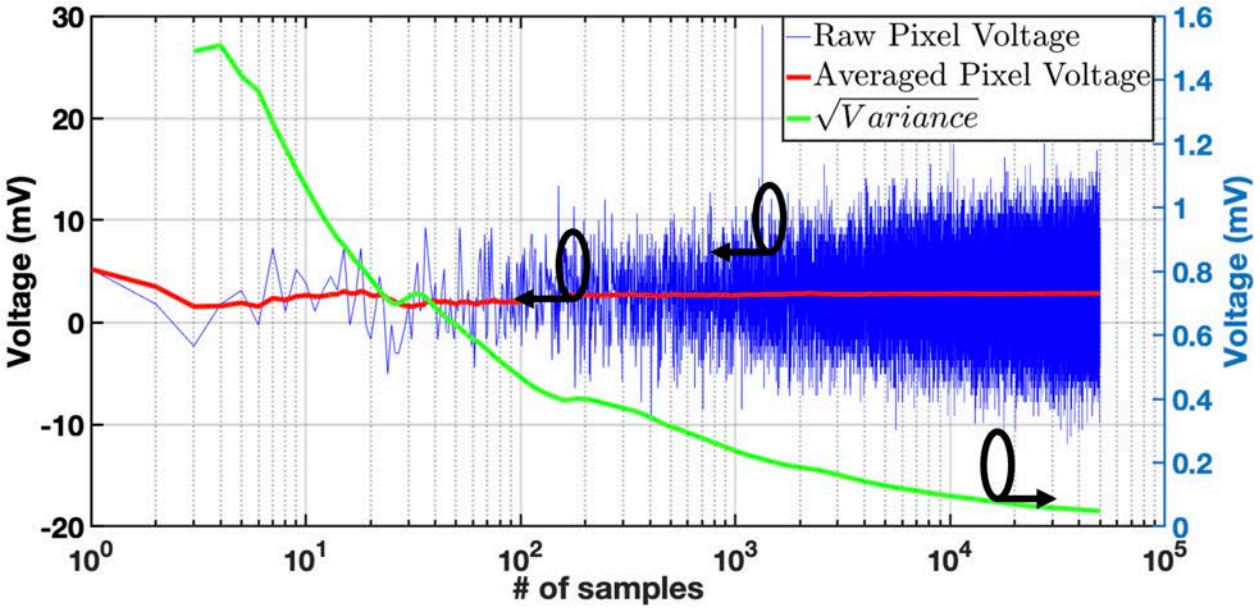


Figure 16. Impact of time-domain averaging on measurement noise and signal reliability.

Measurement noise is critical in obtaining the accurate level of illumination on the photosensor, and the current level of a few mVs of measurement noise is higher than desired and significantly degrades SNR of the pixels. A low SNR results in unreliable pixel values and the variations of the measurements can easily mask the underlying optical signal which can often be lower than the noise power. One way to mitigate this issue is to perform repeated measurements and average the results in time, thus reducing the thermal noise level on the data.

While time-domain averaging may seem trivial, we need to ensure that measurements are repeatable and identical in every other aspect. One key criterion that needs to be maintained consistent throughout the repeated measurements is the signal intensity itself, however, due to photobleaching of most fluorophores and conventional biomarkers this condition is often not satisfied, as the signal intensity degrades and is reduced over time and over repeated excitations. Using a conventional fluorophore therefore does not allow time-domain averaging. Unlike fluorophores, upconverting nanoparticles however, do not suffer from photobleaching and are able to sustain the same amount of emission even after multiple and consecutive excitation cycles, and therefore, their usage will allow us to perform time-domain averaging and help reduce the measurement noise level.

Shown in **Figure 16** is the impact of time-domain averaging on the measurement noise. As seen in **Figure 16**, the raw signal of this sample pixel (shown in blue) has a significant amount of variation, up to +/- 10mV, which can severely limit the reliability of the readout data. However, after >100 averages, this variation is reduced to below 0.4mV rms, which is lower than the variations of the background correction scheme and as a result we no longer become limited by the measurement noise.

It should also be noted that even though averaging many samples helps reduce impact of the noise sources, such as shot noise (the most dominant noise source in this system) and thermal noise, this reduction is only limited to uncorrelated sources such as thermal or shot noise (both being white noise sources). The flicker noise is greatly attenuated by the correlated-double-sampling (CDS) scheme implemented in the pixel, and with integration times in the range of a few milliseconds, the power of the flicker noise residing beyond a few KHz is significantly reduced. However, even though the flicker noise power below a few KHz will remain and no amount of averaging will help eliminate it, this said power is negligible given the very small integrated power that is contained at those frequencies.

4.4 Fully integrated imaging platform

As an intraoperative imaging platform, shown in **Figure 17**, our developed imaging system has at its core a 36-by-80-pixel array, where each pixel includes a photodiode pair (NPD and PPD) and column buffers for time-domain duplexed readout. The sensor also includes a digital control block for readout and acquisition control as well as a row decoder paired with column-wise multiplexers to enable parallel readout of 80 columns using only 8 channels. The acquisition and readout timings are controlled by an FPGA, using digital control signals, and readout data is sent to the FPGA with 8 external 12-bit analog-to-digital converters (ADCs), serially communicating with it. The readout process is carried out by the pair of data lines labeled “SIGNAL”, representing the main CDS sample, and “RESET”, representing the baseline sample for the CDS scheme.

The imaging sensor also includes a current calibration scheme, implemented to correct integration capacitor variation throughout the array, which translates to a pixel gain error, and can be limiting factor for resolving images with low emission light levels. This calibration is an initial one-time step and is performed by directly connecting a single and constant current source to each

pixel successively, extracting the current-to-voltage gain of each pixel relative to each other and using those stored values to adjust the gain during sample acquisition accordingly. **Figure 17** shows how the singular calibration current source is multiplexed and connected to each pixel.

Using direct contact imaging, the specimen will be placed directly on the chip, as shown

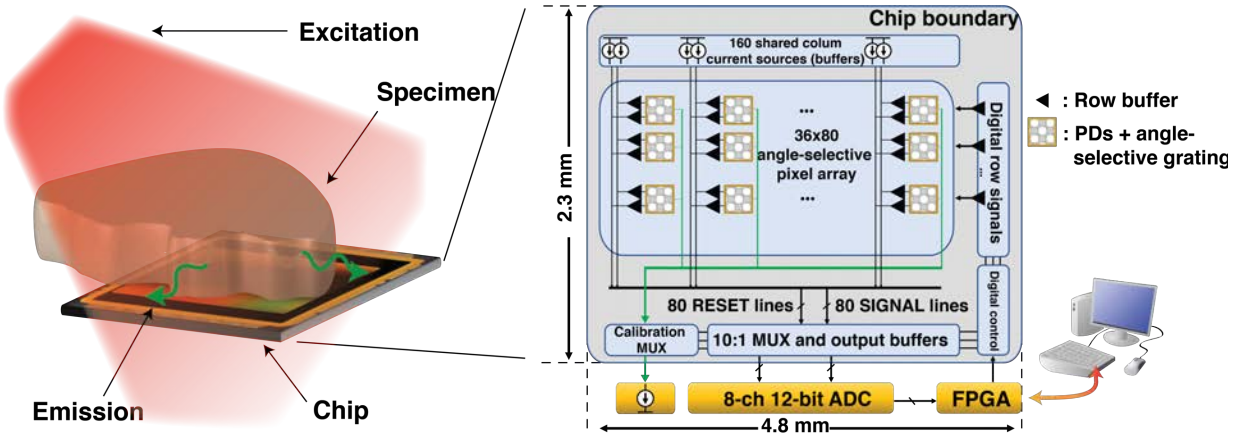


Figure 17. System level overview of the imaging platform, including the specimen, external controlling FPGA, and image sensor and its internal architecture.

in **Figure 17** while the pulsed excitation light is provided by the external light source. The decaying emission light will subsequently be captured and acquired by the sensor using a time-resolved acquisition scheme.

The chip microphotograph of the imager array, fabricated in a 0.18 um process, is shown in **Figure *18***(a). The sensor measures 2.3 mm by 4.8 mm and includes internal digital control blocks to monitor and carry out the acquisition and readout of the array via an external FPGA. **Figure *18***(b) shows a microphotograph of the pixels. As seen in **Figure *18***(b), the negative PD is covered by 5 metal layers and the 4 identical PPD are connected in parallel and covered by angle-selective-gratings, surrounded by control and readout circuitry of the pixel. The 55 µm by 55 µm pixel has been designed to allow efficient use of photodiode areas and shapes to minimize fill factor loss, while accommodating the photodiode pair in it.

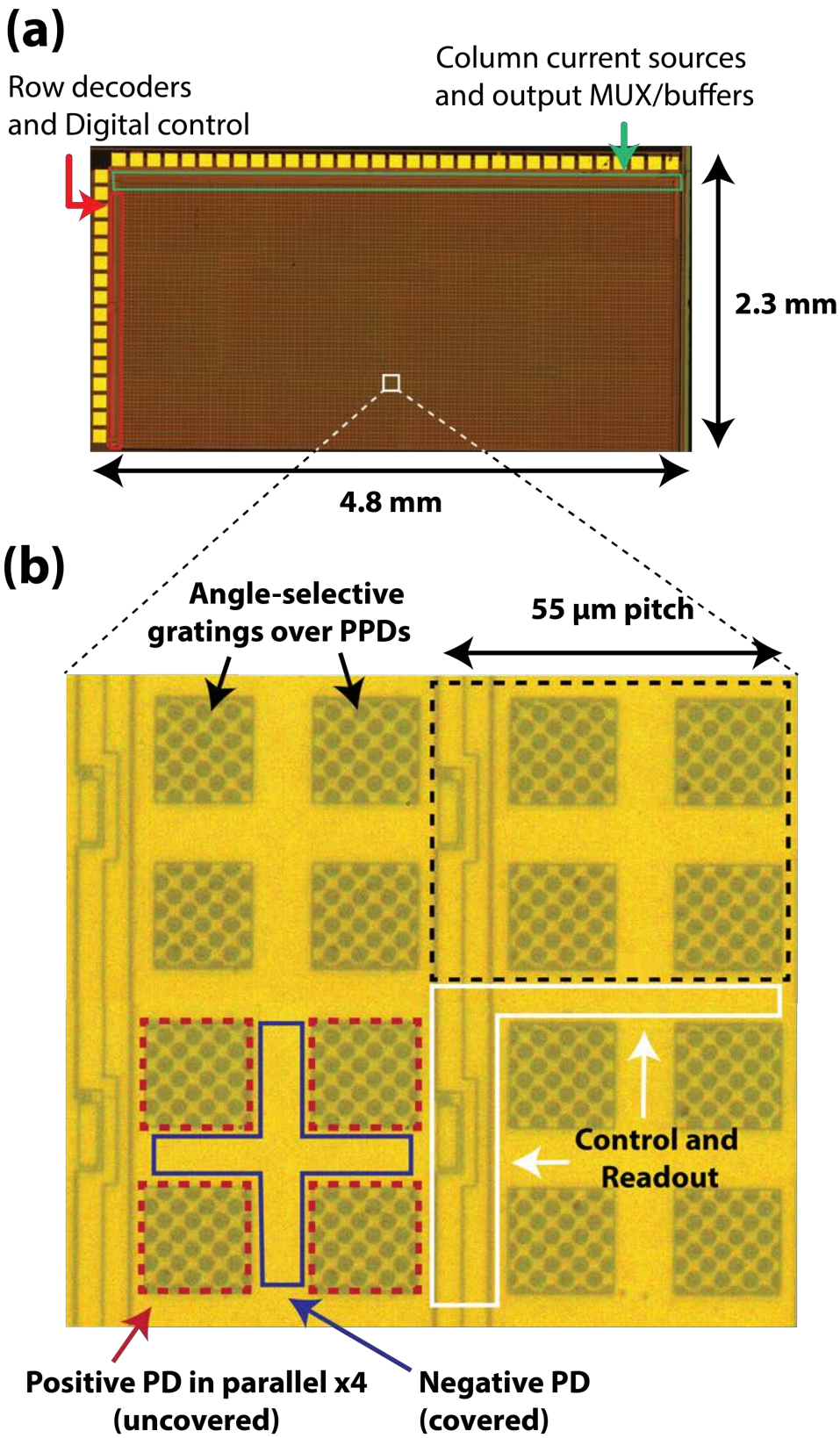


Figure 18. Chip microphotograph: (a) pixel array and digital and control blocks. (b) pixel microphotograph close-up.

4.5 Integration capacitor calibration

Aside from noise, dark current and optical signal variation, another element that was investigated in this work was the variation of the integration capacitors. The integration capacitors being custom-made MOM caps, their variations and mismatch cannot be reliably simulated in CAD using available tools and data and as a result a calibration algorithm was designed and implemented to correct for mismatches that will occur during fabrication.

The mismatch between integration capacitors of different pixels is directly related to the minimum detectable signal and can be a limiting factor in determining the lowest possible signal that can be measured with enough reliability. In a perfectly matched array, the gain of the front-end amplifier will be the same for all pixels and as a result, any difference observed in the levels of the pixels at the output is a reflection of an optical signal difference being measured by the corresponding photodiodes. However, in the presence of mismatched integration capacitors, and therefore mismatched TIA gains, a difference in the measured pixel levels does not necessarily carry optical information since part of this level difference is a direct result of the different conversion gain of the pixel themselves.

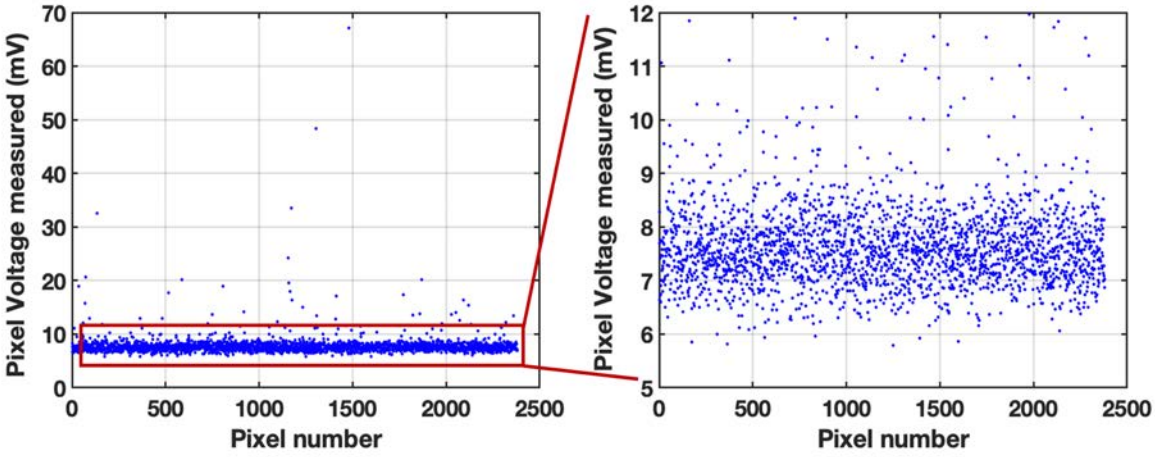


Figure 19. Measured pixel levels with only dark current as input (at integration time of 1ms).

To calibrate the mismatch between conversion gains of the pixels, one needs to extract the relative gains of the pixels, or effectively, measure the relative mismatch between the integration capacitors of the pixels.

One simple method to perform this measurement is to measure the conversion gain of every pixel using one of the current-type sources already available in the pixel, namely the photodiode itself. However, care must be taken to ensure that the signal is exactly uniform across all pixels, which would require a perfectly uniform external optical source illumination the sensor – which is not realistic of even feasible given the precision needed. Another source that could serve as the input is the dark current of the photodiode themselves, and while this method would no longer necessitate an external light source, it still presents a very non-uniform profile throughout the array and is inherently heavily stochastically random across the pixels – not useful for calibration. **Figure *19*** shows the measured pixel levels at 1ms of integration time, where the spread of the levels is mainly due to the randomness of the dark current being the sole input in this experiment. Using

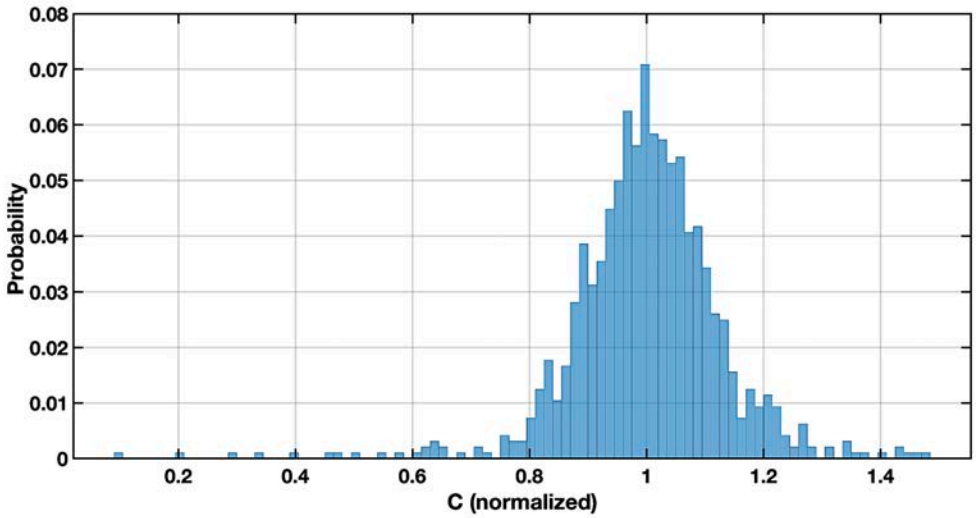


Figure 20. Distribution of measured relative weights of integration capacitors using only dark current as input source.

the data in **Figure *19***, we can extract the relative values of the integration capacitors, which are shown in **Figure *20*** where every pixel was calibrated using its own photosensor's dark current. As observed in **Figure *20***, the variation of the measured capacitor is significant however the majority of this spread is a result of the randomness of the dark current itself which dominates the MOM cap variation in this experiment.

A more accurate way of performing this calibration is to use a single and unique current source as input to all pixels. To this end, an additional input path was added to every pixel and routed to a single and unique current source set by an external reference. During calibration, this current source will be connected to every pixel, one at a time, and the conversion gain of the pixel will be measured accordingly. Since the input to all pixels will now be identical, the resulting variation and spread of the conversion gain is only corresponding to the capacitor variation, and can therefore be used to correct gain errors during regular data acquisition. A diagram of the implemented scheme is shown in **Figure *21*** illustrating the use of a single current source to stimulate every single pixel in the array.

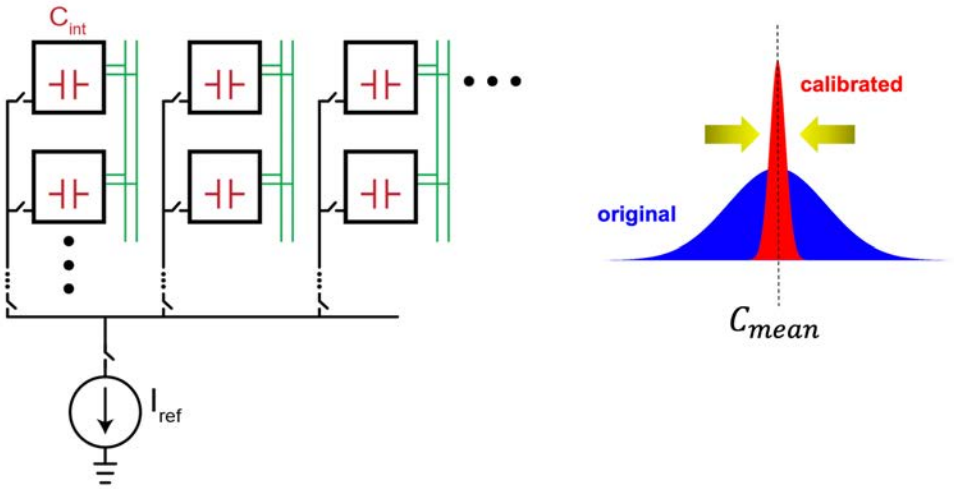


Figure 21. Calibration scheme for integration capacitor using a single and unique current source, to correct gain errors and capacitor stochastic randomness.

Using an external current source of 0.55uA and after several current attenuation, the attenuated current is connected to each pixel and the pixel levels were measured at various integration times and the resulting measurement are shown in **Figure *22***. To measure the relative conversion gains, the gains were calculated using two integration datapoints (0.1 ms and 0.6 ms) and the resulting histogram of the gains are extracted and shown in **Figure *23***.

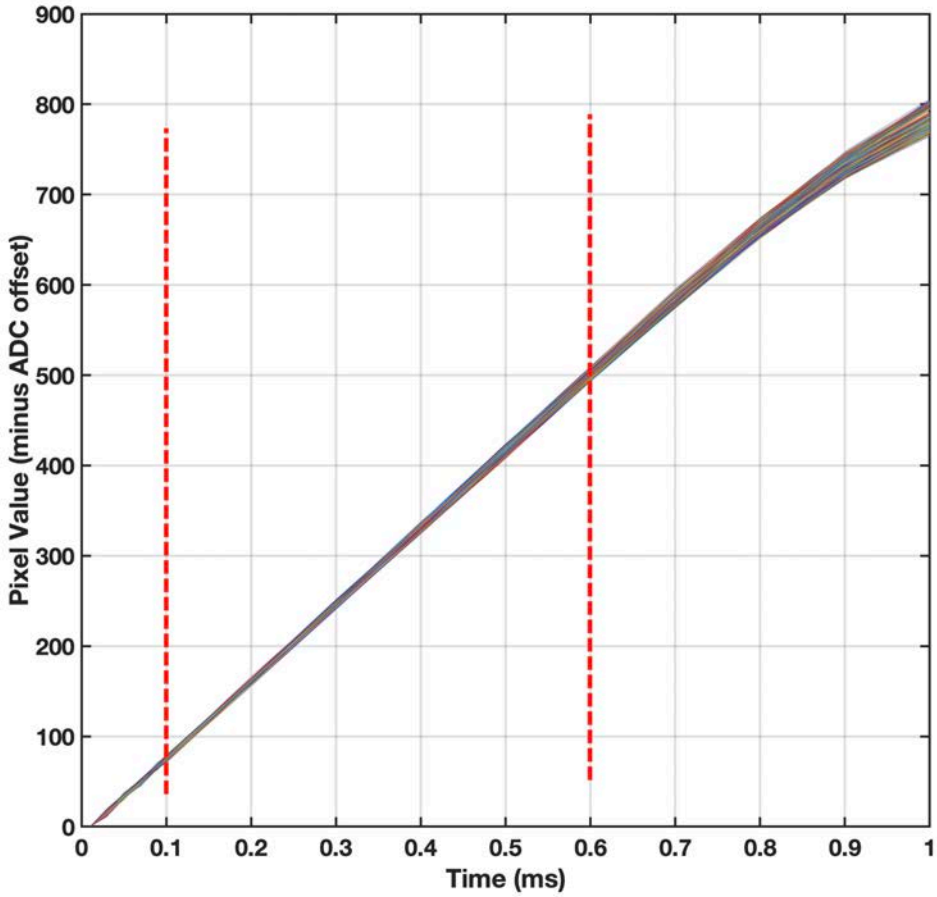


Figure 22. Pixel levels using calibration current source measured at various integration times.

The gain histogram in **Figure *23*** shows that the distribution of the capacitor values is quite narrow and does not currently impose a limitation on the overall system performance.

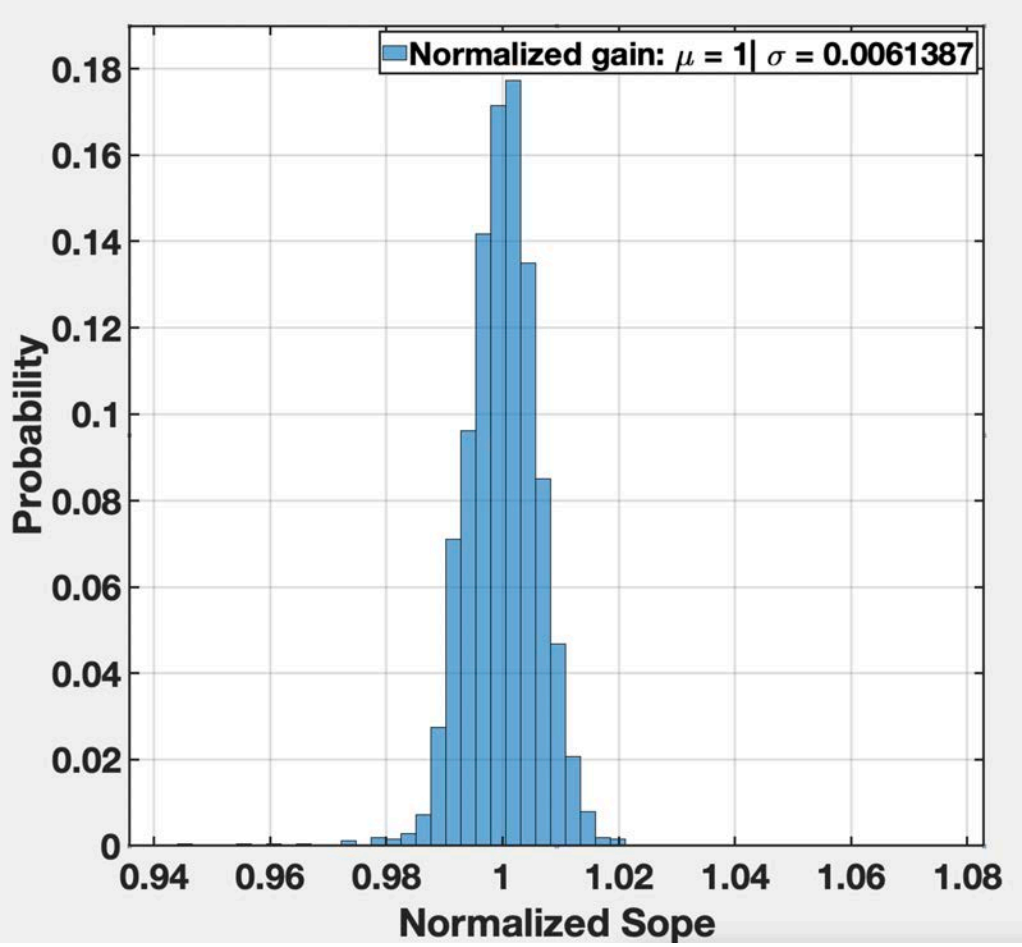


Figure 23. Extracted histogram of conversion gain of the pixels during calibration.

4.6 Impact of die thickness

As described before, the main limiting factor in optics-free time-resolved imaging is the introduction of optically-induced carriers in the substrate that creates an interference on the photosensor, when they inevitably reach the active region in order to recombine and dissipate, and as a consequence, mask or corrupt the desired optical signal captured by the photodiodes. The thickness of the substrate plays a critical role in the creation and dissipation of these charges. Since the NIR light is more penetrative into Silicon than visible or UV light, the larger the volume of the silicon in the beam's path is, the more charges and electro-hole pairs will be created throughout it. Added to this is the fact that these optically-induced charges have a very long lifetime (>1ms) – thanks to the low doping in the substrate – and the combination of these two properties provides these charges with the perfect condition and sufficient time to reach the active region and recombine at the photosensor sites. However, the thinner the substrate is, the smaller the total amount of charges created will be and in addition, the shorter their lifetime will become since the path they need to travel to the surface and recombine is now much shorter.

While the magnitude of these charges can depend on many other factors such as angle of incidence, wavelength and metal structures obstructing the beam, the lifetime of these interference generated on the photosensors however is directly dependent on the thickness of the substrate, and the smaller the thickness of die is the shorter this lifetime will be. To verify this, 4 identical dies were thinned down to 4 different thicknesses and the lifetime of the relative interference generated on each of them under similar conditions was measured and the results are shown in **Figure *24***. The thickest die (200 micron) still showed a significant amount of interference on the pixel even after 1ms whereas the 25-micron thin die was free of any interference at 500 us, showing the direct correlation between die thickness and lifetime of the substrate carriers.

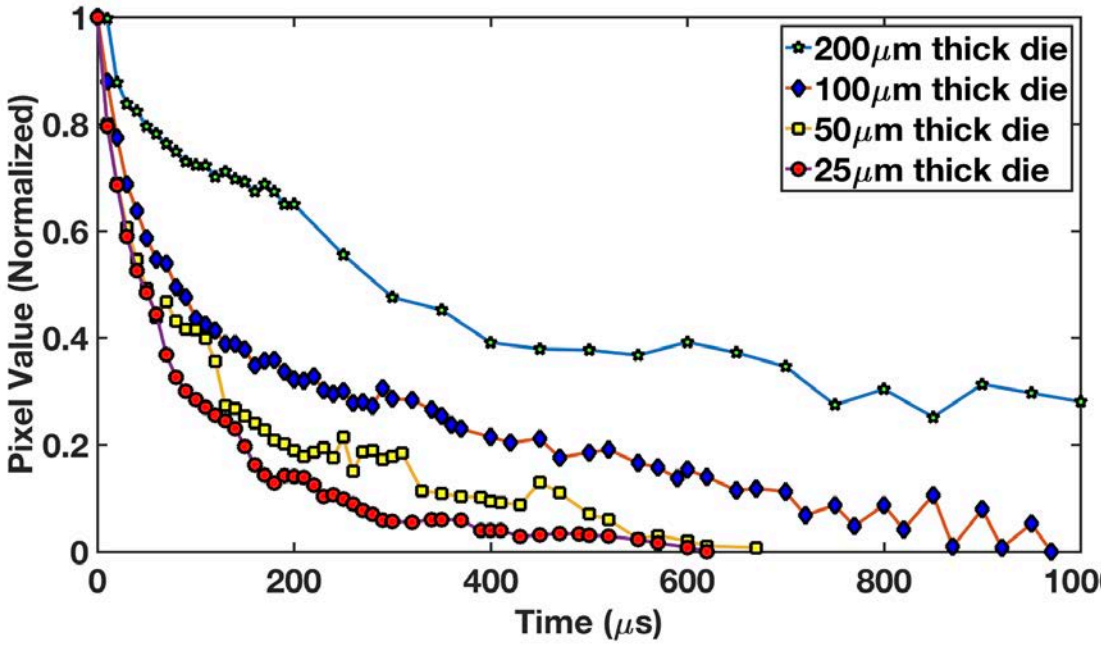


Figure 24. Impact of die and substrate thickness on the lifetime of the optically-induced charges.

5 Optics-free specimen NIR imaging

This chapter describes the results of imaging various specimens to determine and quantify the applicability of upconverting nanoparticles in mice injections (i.e., bio-distribution, toxicity, endocytosis, nanoparticle draining from specimen to name a few), the performance of the developed imager array, as well as the background adjustment method described in earlier sections. We have carried out several experiments to achieve a proper characterization of the imager performance, including intratumoral injections of upconverting nanoparticles into HER2+ breast tumors into 6 mice and monitoring biodistribution and toxicity, imaging United-States Air Force resolution (USAF) target plates (negative) in combination with upconverting nanoparticles, extracting spatial resolution and minimum target clearance (distance between two distinct features) needed for the imager to be able to resolve using the USAF line pairs, conducting intratumoral injection into grown prostate tumor in mouse, imaging of this excised tumor specimen as well as imaging of a single-cell layer of this injected tumor sample.

5.1 Bio-distribution

To observe bio-distribution, toxicity, and sentinel lymph node identification, we grew breast HER2+ breast tumors in 6 mice and once mature and ready, we conducted intratumoral injections of the nanoparticles into the mammary fat pads of the mice. The tumors were grown on both sides of the mice, however, to maintain a negative control and reference throughout the experiment, only one side was injected and the other side was used as a negative control. The mice were later individually sacrificed at 6 different time-points and the injected mammary glands were imaged with IVIS spectrum imager (PerkinElmer) with an excitation power of 0.1W/cm^2 (at a wavelength of 980 nm)[21], [29]. **Figure *25*** shows the measured intensity of the emission of the upconverting nanoparticles injected into the mammary fat pad of one of the mice over a period of 6 hours, where we have measured the emission every 2 hours. The intensity of emission measured at 660 nm wavelength (main emission peak of upconverting nanoparticles) was 12.5, 7.6 and 5.8 kcps (counts per second) at the timepoints of 2, 4 and 6 hours respectively. The extracted signal-to-background ratio (SBR) resulting from those measured emission intensities were respectively 44, 29 and 25, showing a decrease of approximately 44% over a span of 6 hours. Despite the intensity decrease observed in **Figure *25***, a significant amount of upconverting nanoparticles still remain in the gland even hours after injection, which is enough time to allow surgeons to conduct a complete guided resection surgery of a breast tumor. These results highlight the compatibility and applicability of upconverting nanoparticles in acting as the molecular marker and optical tag in achieving guided surgery.

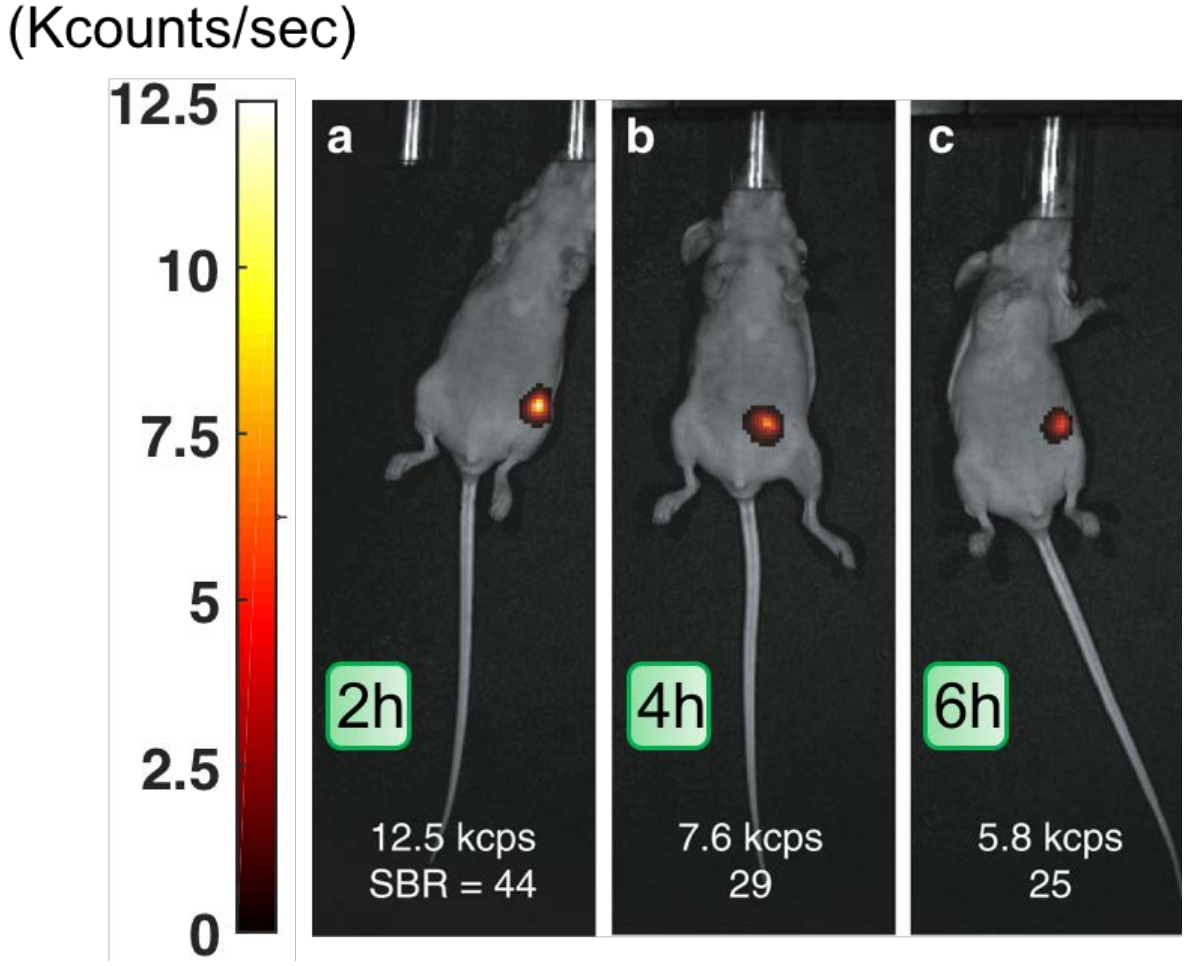
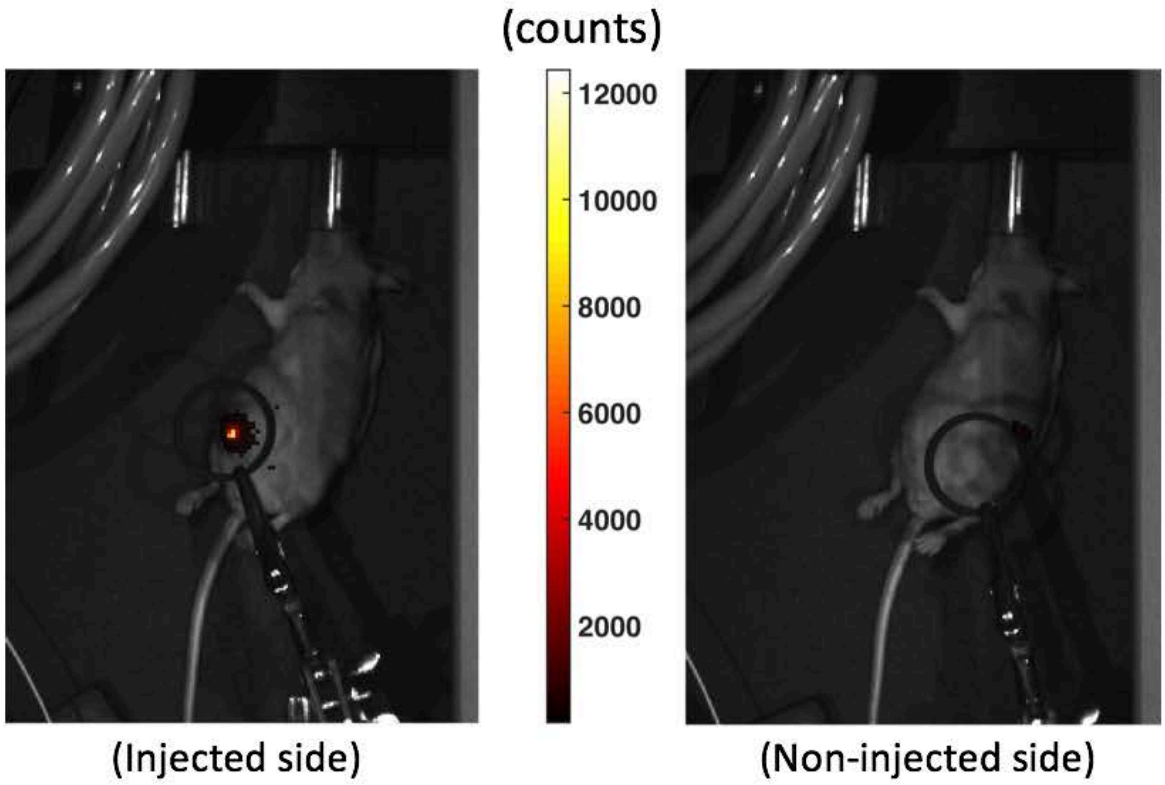
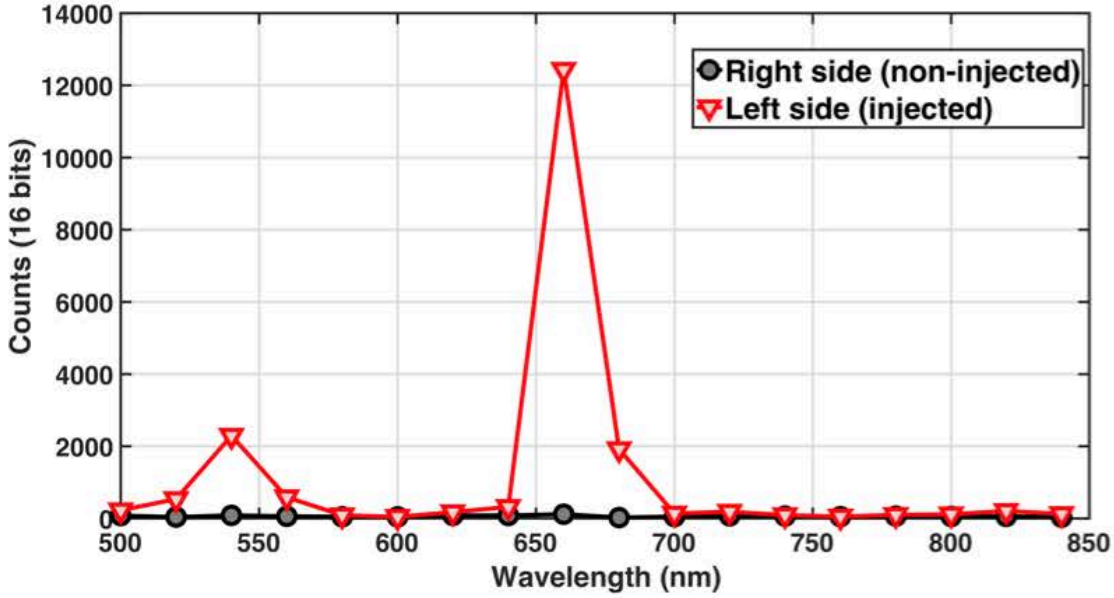


Figure 25. UCNP-injected mammary fat pad emission over 6 hours. (images in counts/sec and obtained using the IVIS imager)

As a negative and reference control, and to ensure that our IVIS images are in fact capturing the emission of the UCNPs and that no autofluorescence or other source of potential interference is being misconstrued for upconverting nanoparticle emission, we have measured the spectrum of the emission from the injected and non-injected mammary fat pad, shown in **Figure *26***(a), and the resulting spectrums are illustrated in **Figure *26***(b). Knowing that the UCNPs mainly emit in the 550 nm and 660 nm range, based on the spectrum obtained, we can assert that the signal being observed is in fact from the UCNP and surrounding tissue and cells do not generate any source of additional background or autofluorescence on the IVIS imager.



(a)



(b)

Figure 26. UCNP-injected mammary fat pad emission spectrum (images obtained using the IVIS imager): (a) left side (injected) and right side (non-injected) IVIS images (obtained under 0.1 W/cm^2 continuous 980 nm light). (b) left and right side emission spectrum (obtained under 0.1 W/cm^2 continuous 980 nm light).

5.2 USAF resolution target plate imaging

To validate the ability of our sensor to perform time-resolved imaging using upconverting nanoparticles, we tested its performance in a specimen experiment. **Figure *27***(a) shows the

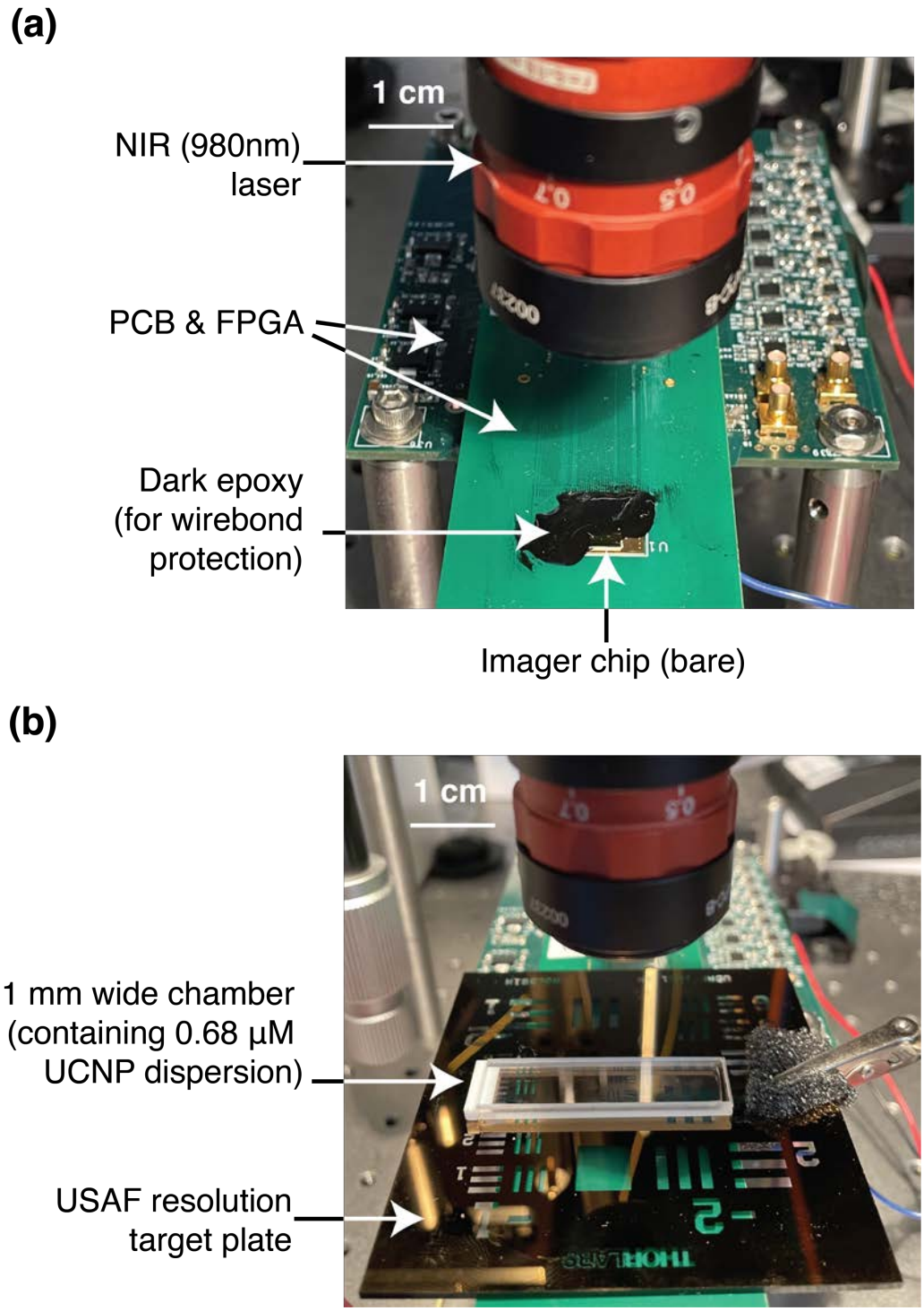


Figure 27. Overview of the setup used for imaging: (a) before mounting the specimen on the imager. (b) after mounting the specimen on the imager.

imaging platform (without the specimen mounted). Due to our ability to perform direct contact imaging, the platform does not include any filters or lenses, and as a result, the imager chip can be used bare, as shown in **Figure *27***(a).

The specimen used for the experiment is a negative USAF resolution target test plate, coated with a chrome background. We have selected a singular feature on the plate -the digit "1" in the "group 1" section of the plate- and placed the plate directly on the sensor. This specimen is illuminated by a quartz chamber containing upconverting nanoparticles dispersion, which is illuminated by an external excitation light source. 500 µL of a 0.68 µM upconverting nanoparticles dispersion (in hexane) is transferred into the quartz chamber, which has a 1 mm light path (thickness), placed directly on the USAF target plate, as shown in **Figure *27***(b). Once the specimen is mounted, the laser is collimated and pulsed, and the time-gated acquisition of the specimen emission is carried out. The laser pulse duration and integration time were respectively 5 ms and 1 ms, resulting in an overall frame rate of 34 Hz. The laser power used for this experiment was 200 W/cm^2. This experiment has been designed to emulate a tumor specimen that has been injected with targeted upconverting nanoparticles, where the imaging platform will serve as a surgical tool for clinicians to direct and guide them to locate and visualize tumor bed layers and microscopic residual cancer cells being left behind during resection surgeries.

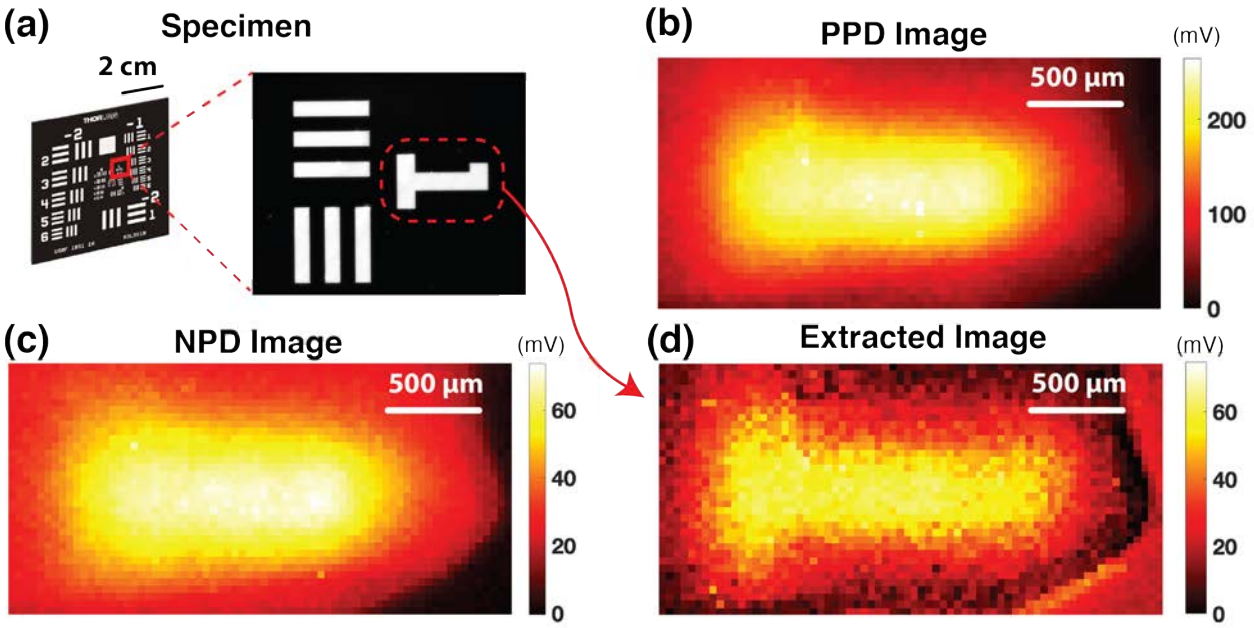


Figure 28. Specimen experiment results: (a) image of the feature on the USAF resolution target plate being acquired. (b) measured time-resolved image of the positive photodiode (T_{int} = 1 ms). (c) measured time-resolved image of the negative photodiode (T_{int} = 1 ms). (d) extracted and background-corrected image of the specimen.

The ground truth image of the specimen (digit "1") is shown in **Figure *28***(a). The background correction method relies on the values extracted from the positive PD and the negative PD, both of which are shown in **Figure *28***(b) and **Figure *28***(c) respectively. The PPD measured an average of 225 mV of signal where emission was present, while the average NPD's measured signal in that same region was 62 mV, both of which include an emission component as well as the background from the NIR excitation. Using the background adjustment method proposed, we can extract the underlying emission component of the image and resolve the image shown in **Figure *28***(d). The average emission signal in the final extracted image is 50 mV, for an integration

time of 1 ms. The background in **Figure *28***(d) (surrounding the specimen) is less than 10 mV and has a mean of 6.5 mV, demonstrating that the proposed method is able to reduce the interfering NIR excitation to single-digit mV level and near the noise floor of 2.2 mV, thus allowing filter-

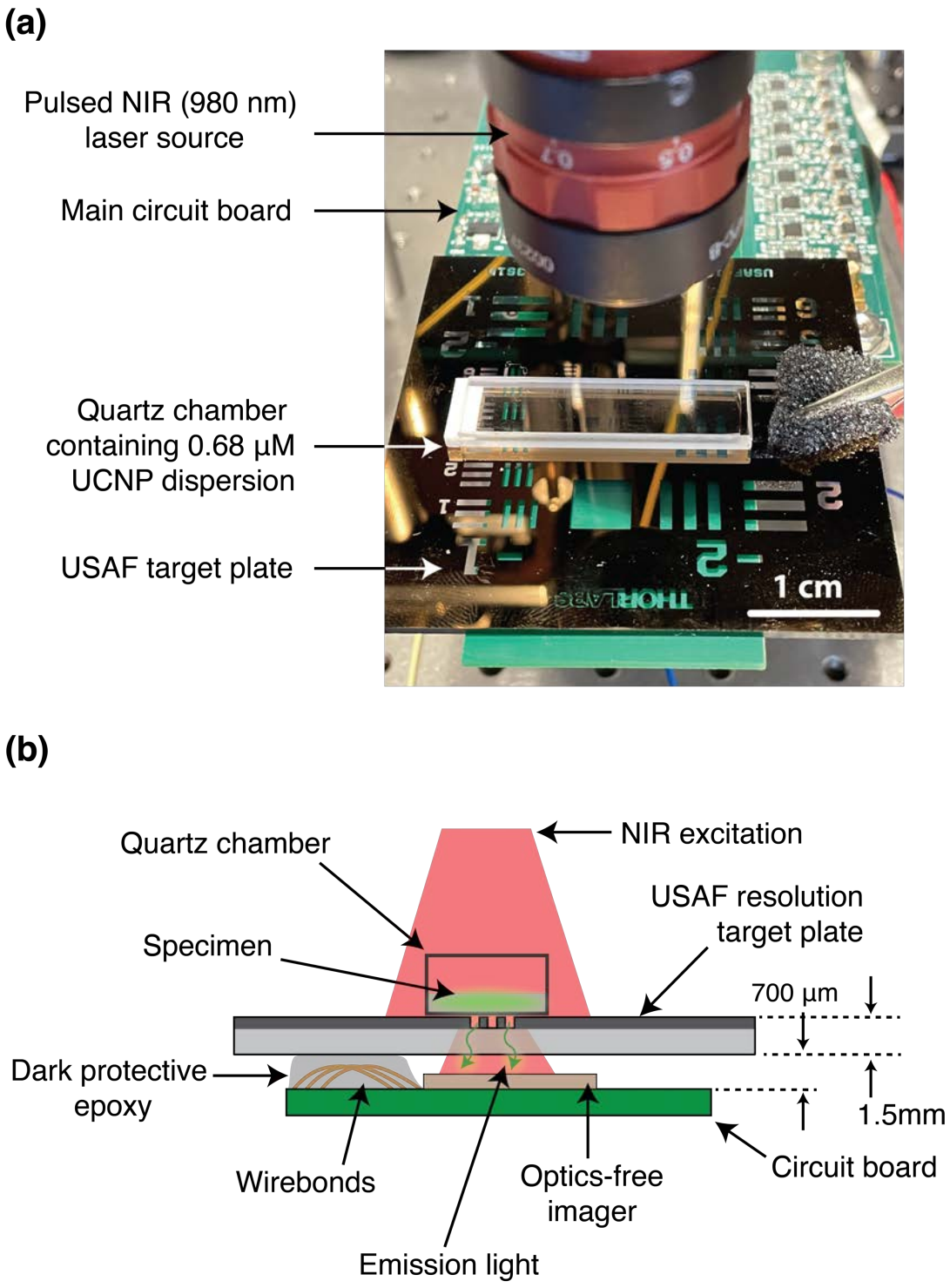


Figure 29. Specimen experiment results: (a) image of the feature on the USAF resolution target plate being acquired. (b) measured time-resolved image of the positive photodiode (T_{int} = 1 ms). (c) measured time-resolved image of the negative photodiode (T_{int} = 1 ms). (d) extracted and background-corrected image of the specimen.

less and lens-less time-resolved imaging to be possible. The angle-sensitivity of the background adjustment method remains the main limiting factor to achieve higher sensitivity, surpassing integration capacitor and photodiode responsivity mismatch between the pixels, each only resulting in at most 1.5 mV of measurement error.

The next experiment with the USAF resolution target plate was carried out to determine the spatial resolution of the micro-imager, for which we used the negative resolution target plate to extract the smallest spatial feature this imaging system would be able to properly resolve. The setup used for this experiment is shown in **Figure *29***(a), where the USAF resolution target plate is placed directly on the imager, and a quartz optical chamber (with 1 mm optical path) containing 0.68 µM of UCNPs dispersion is placed atop the plate as a source of emission for the purposes of this experiment. A collimated 980 nm laser, emitting 5-ms long pulses of 18 W/cm^2 illuminates the entire system, including the micro-chip and the specimen. The laser is controlled by an external controller module tasked with maintaining proper timing and synchronization between excitation pulses and other relevant control signals. The distance between the emission source (the UCNP chamber) and the micro-imager is 2.2 mm, 700 µm of which are necessary for protecting the

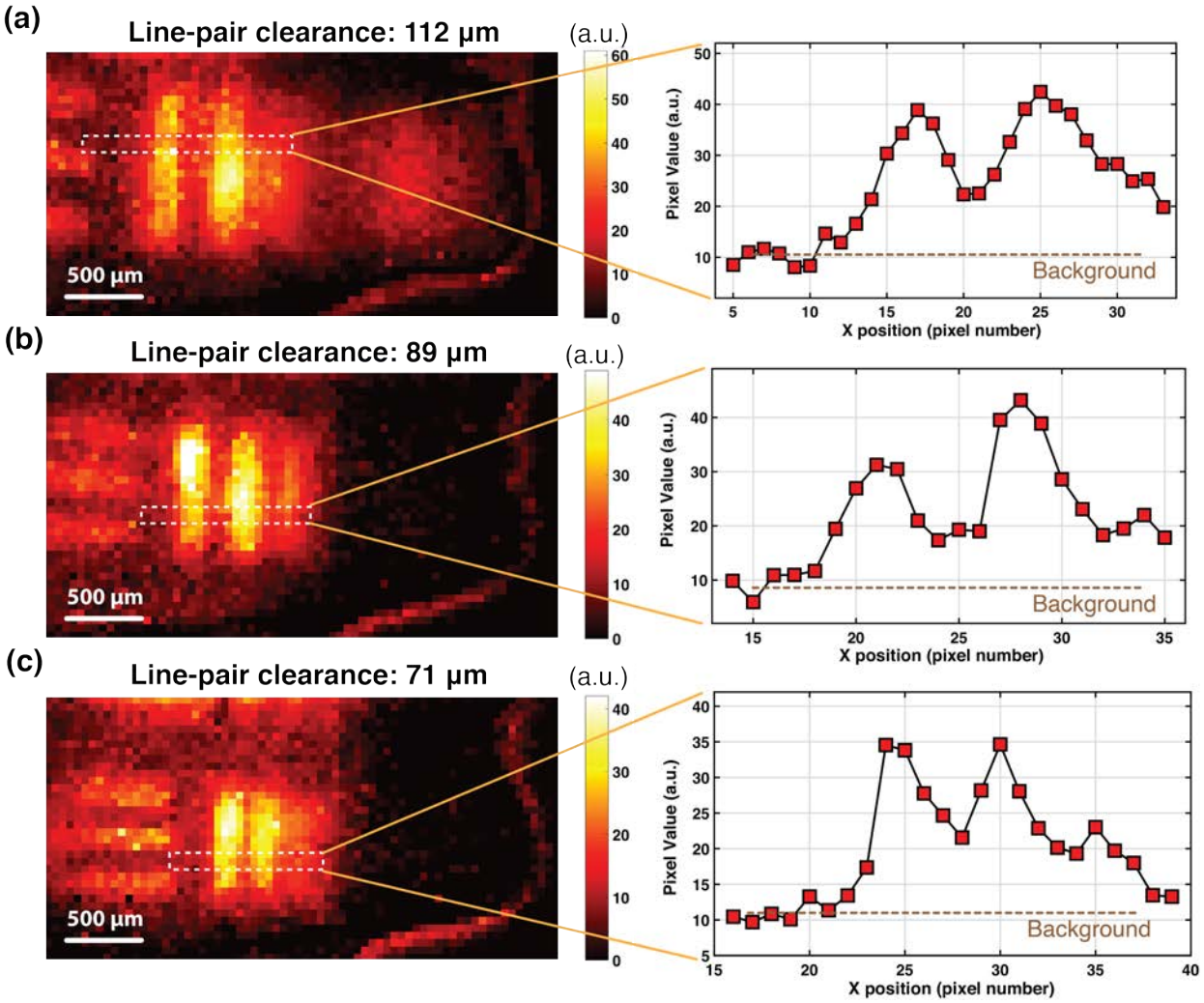


Figure 30. Acquired images and signal intensity cross section of 3 line-pair clearances (distance between line pairs) on the USAF resolution target plate (illuminated with 5-ms long pulses of 18 W/cm^2 980 nm light): (a) line-pair clearance of 112 µm. (b) line-pair clearance of 89 µm. (c) line-pair clearance of 71 µm.

surrounding wirebonds of the imager from mechanical stress, as shown in **Figure *29***(b). This separation margin can be virtually eliminated if a more compact electrical contact interface is used for the sensor -e.g., ball-grid array (BGA) contact-pads.

We have imaged 3 line-pairs on the USAF resolution target plate with clearances of 112 µm, 89 µm and 71 µm respectively, during which localized background adjustment was performed upon acquiring the image. To determine whether a line-pair has been successfully resolved, we defined "half-width" as the range of pixels in which the amplitude of the emission is more than half the highest intensity, all measured relative to the background level. This metric has been used to determine the spatial resolution of the sensor in this experiment.

The final resulting images are shown in **Figure 30**. **Figure 30**(c) illustrates the case for the line-pair with the smallest clearance (71 µm), with a background of 10 a.u., and a value of 35 a.u. for highest intensity. The half-width range is therefore limited to pixels higher than 22.5 a.u., resulting in only a single pixel below the half-width limit (= pixel in position 28).

A singular pixel distinction between two neighboring spatial features –line-pairs in this case– is the practical limit for spatial resolution, and as a result, the smallest spatial target our proposed imager platform is able to resolve is 71 µm. This minimum achievable resolution is largely limited by the accuracy of the background adjustment scheme, and the blocking performance of the integrated in-pixel micro-collimators[23].

While our experimental results with USAF target plate demonstrated enough resolution to be considered a reliable imaging tool for intraoperative imaging, its performance can be further improved if clinical and surgical settings allowed increasing the form factor of the imager. The

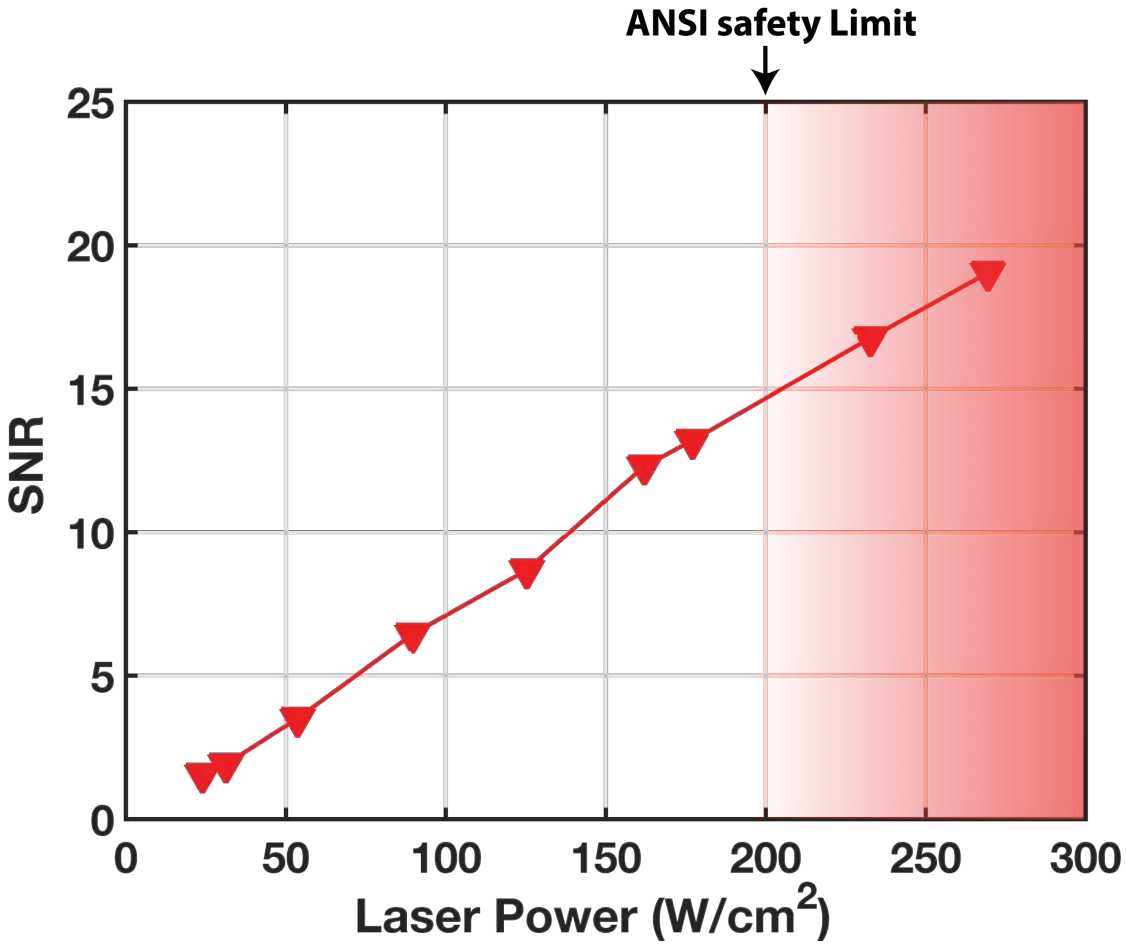


Figure 31. SNR of pixel (in emission region) as a function of excitation light power.

sharpness and resolution of the imager can be further enhanced by further increasing the aspect ratio of the angle-selective micro-gratings. If sufficiently thin, they would provide only negligible increase in form factor, but significantly increase the spatial resolution of the imager -at the expense of a reduction in signal strength.

As an intraoperative imager, our platform must also ensure full compliance with clinical optical and maximum exposure limits. According to the American National Standards Institute (ANSI) the limit for a 5 ms long pulsed laser of 980 nm, the maximum allowed power density is 200 W/cm^2[32]. **Figure *31*** shows the SNR change with excitation power density, and as seen in the graph, despite the ANSI limit of 200 W/cm^2, the imager is still able to achieve a maximum SNR of 15dB.

5.3 Imaging intratumorally injected UCNPs

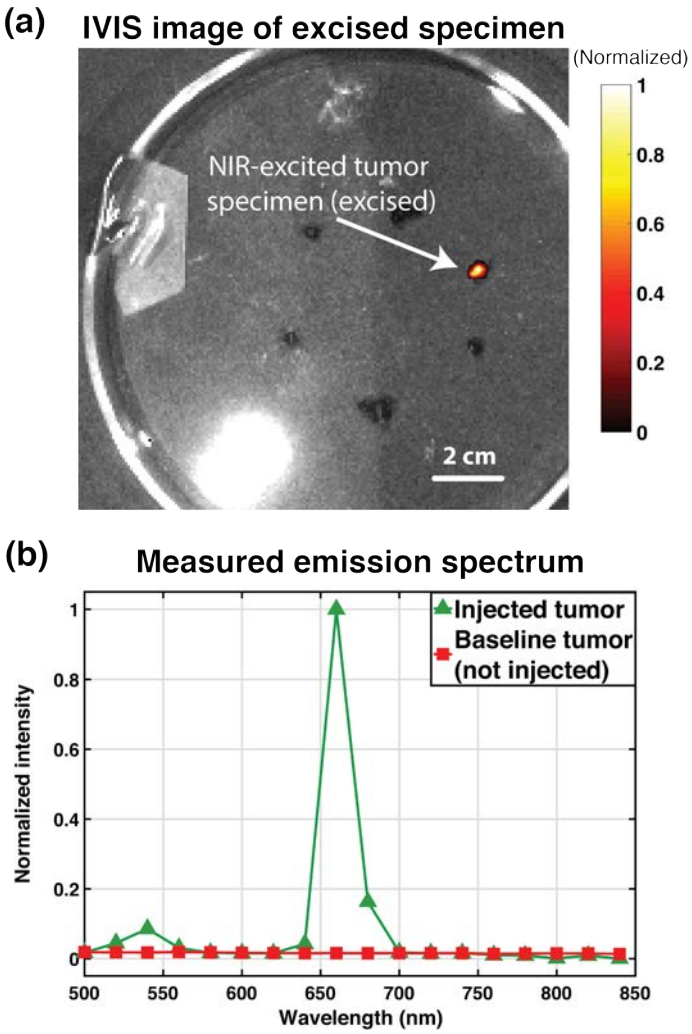


Figure 32. Excised prostate tumor imaging results with IVIS spectrum imager: (a) image of excised tumor specimen acquired using the 660 nm emission filter (with 20 nm pass band) on the IVIS spectrum imager (excitation provided by custom-modified and external continuous 980 nm laser source). (b) measured emission of specimen under continuous 980 nm excitation light (22 W/cm^2).

In order to demonstrate the performance of our sensor in imaging tumor samples, we injected a prostate tumor with an aqueous solution of polymer-encapsulated upconverting nanoparticles (with a concentration of 250 nM) and subsequently excised that tumor for imaging (see [33] for more details on the sample used). These upconverting nanoparticles were synthesized using a 16 nm $NaEr_{0.8}Yb_{0.2}F_4$ core and a 5 nm $NaY_{0.8}Gd_{0.2}F_4$ shell [21] and the injection volume was 25 µL. The excised tumor was imaged first using a custom-modified IVIS imager equipped with an NIR-I illumination source –980 nm wavelength stabilized single mode fiber coupled laser diode– and additional and necessary rejection filters, in order to verify the colocalization of the specimen and extract the spectrum of the UCNPs emission within the sample. The laser beam size being limited to only a few millimeters in diameter, only areas of interest are illuminated (excited) in this experiment. All emission spectra were extracted under a continuous 45 W/cm^2 980 nm illumination.

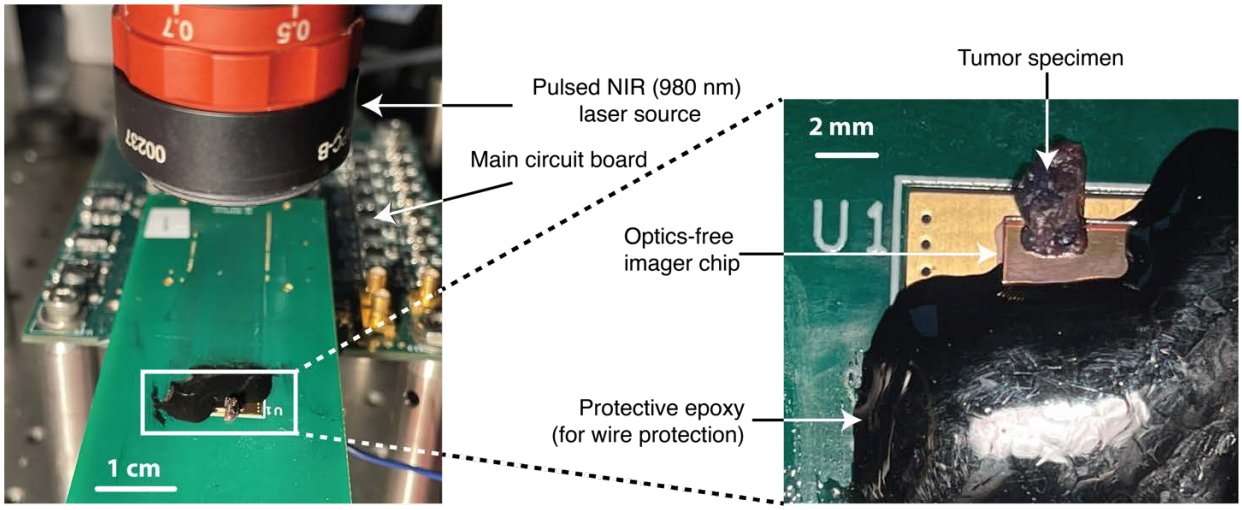


Figure 33. Experiment lab setup for *ex vivo* imaging of intratumorally-injected prostate tumor with alloyed upconverting nanoparticles (illuminated with 5-ms long pulses of 45 W/cm^2 980 nm light).

Figure *32*(a) shows the excised tumor imaged on the IVIS imager and the measured emission of the specimen is plotted in **Figure *32***(b). The emission spectrum of a tumor without UCNPs (under 980 nm excitation) is also shown in **Figure *32***(b) as baseline. The emission measured displays the two clear major visible bands of emission of the UCNPs at 545 nm and 655 nm. The intensity of the measured emission at 545 nm was 8.5% lower than the level at 655 nm, while the measured SBR in **Figure *32***(a) was 33.5.

Upon extracting the emission spectrum, the specimen was then prepared to be imaged on the custom-fabricated micro-imager, as shown in **Figure 33**. The specimen is directly placed on the imager chip, with no external optical components (i.e., filters and lenses).

The micro-chip is also supported by two connected electronic circuit boards and controlled by an external computer to ensure timing and synchronization is maintained throughout the acquisition process. The controlling computer system is also tasked with communicating with the micro-imager and retrieve and read out acquired images and perform localized background adjustment after acquisition. Additionally, the excitation laser timing is digitally controlled by this control unit to generate pulses of 45 W/cm^2 NIR light each lasting 5 ms, synchronized with the rest of the acquisition process.

The tumor specimen was imaged with the optics-free custom-fabricated micro-imager using the time-resolved imaging method described in earlier sections. Shown in **Figure *34***(a) is a high-resolution microscope image of the tumor specimen, acquired as a ground truth. To ensure a fair quantitative comparison of the ground truth image and the micro-imager result, the microscope image has been scaled to match the pixel size of the custom-fabricated imager, and the resulting image is shown in **Figure *34***(b).

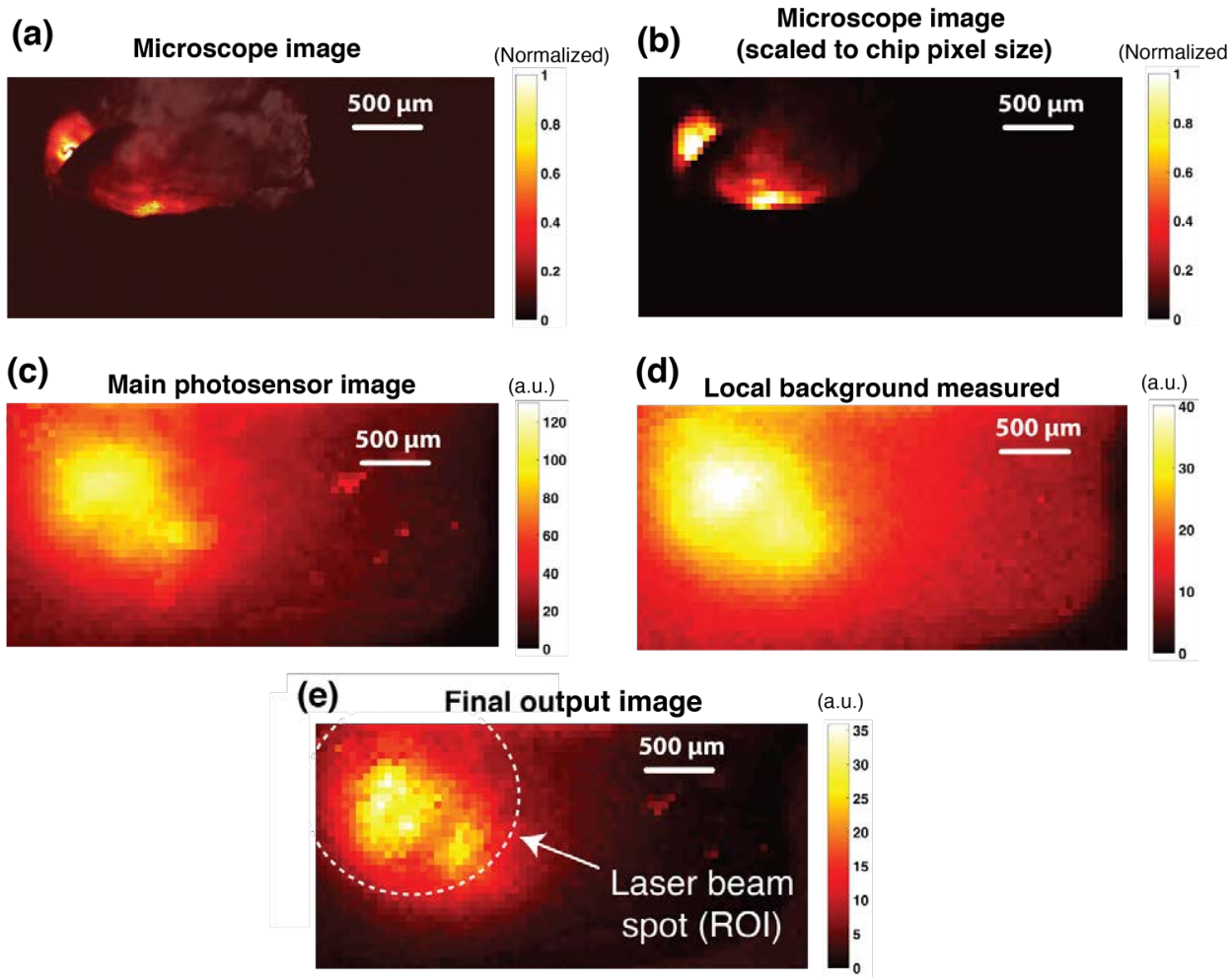


Figure 34. Ex vivo experiment images: (a) high-resolution microscope image of excised intratumorally injected prostate tumor (T_{int} = 1s). (b) microscope image of excised prostate tumor with matched pixel pitch (to micro-chip sensor). (c) main photosensor image of micro-chip sensor. (d) secondary photosensor image of micro-chip sensor. (e) image of main photosensor after applying the background adjustment and correction scheme.

The image captured using the main photosensors in each pixel is shown in **Figure *34***(c) displaying a significantly high level of background, virtually masking the underlying emission signal from the specimen, highlighting one of the most critical limitations of prior works such as [42, 44]. To correct this background level, an additional acquisition is performed to capture the data from the secondary (fully covered) photosensors and extract the locally measured background (**Figure *34***(d)). These measured local background levels are then used to adjust the background on the main photosensor image and recover the underlying emission of the specimen. **Figure *34***(e)

shows the main photosensor image after background correction, revealing the emission signal, with maximum measured signal of 35 a.u. and an SBR of approximately 8. This new dual-photosensor architecture was able to correct more than 80 a.u. of background measured in **Figure *34***(c) –where the effective SBR was 35/80 or 0.4, enhancing the SBR by 20x, with no additional optics or increase in form factor. **Figure *34***(e) reveals the correlation of the micro-chip image with the microscope ground truth data shown in **Figure *34***(b), demonstrating the capability of the silicon-based sensor in becoming a significantly smaller and surgically practical alternative for high resolution cumbersome microscopes.

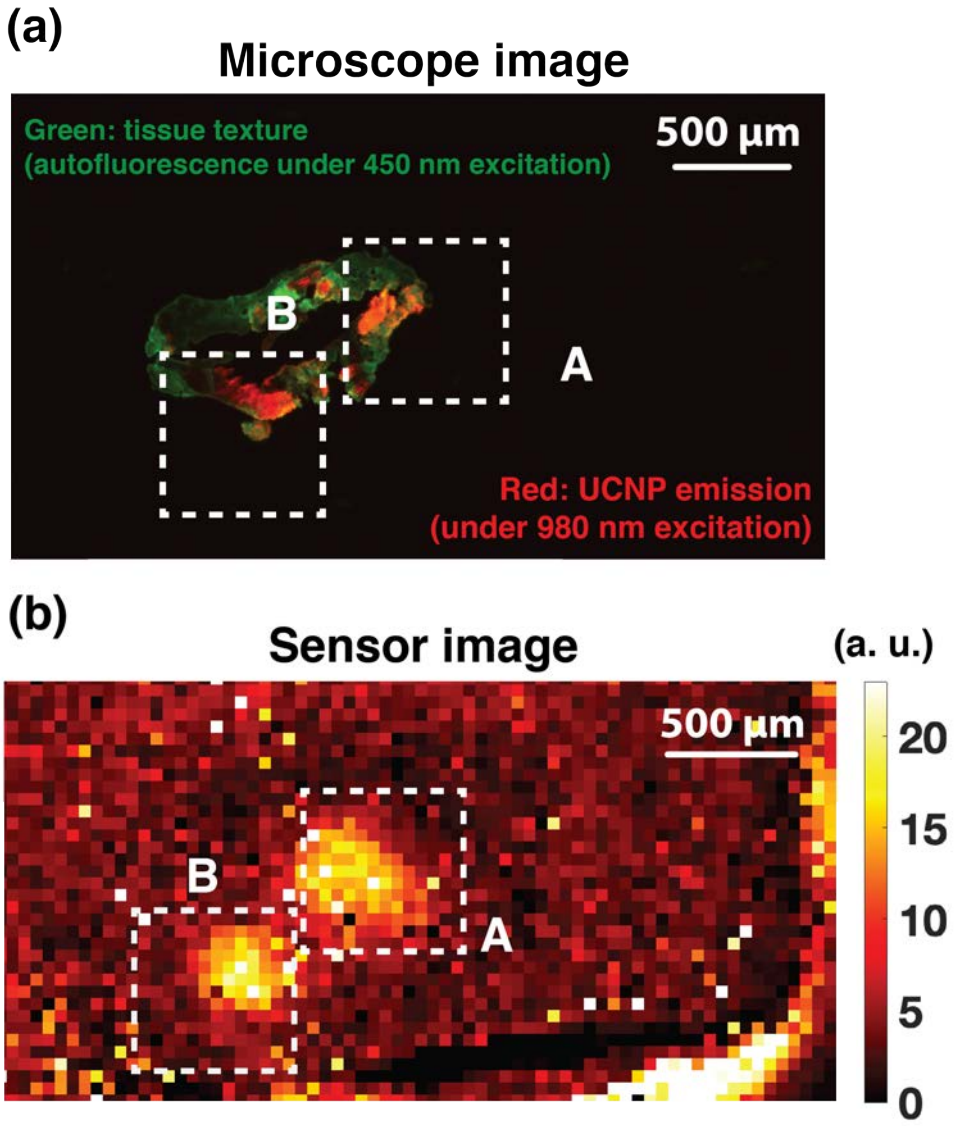


Figure 35. Single-cell thin layer imaging experiment: (a) composite microscope image of a 14 µm thin section of intratumorally injected prostate tumor specimen, with the green color representing tissue texture obtained using autofluorescence under a 450 nm excitation light, and red representing the UCNP emission captured using a continuous 45 W/cm^2 980 nm excitation (Tint = 1s). (b) background-adjusted image of the 14 µm thin section captured on the micro-chip under pulsed 980 nm excitation (45 W/cm^2).

As part of our last experiment, the tumor sample was later sectioned into 14 μm thin (single-cell layer) sections placed on glass slides and the resolution and performance of the imaging sensor was evaluated on a sample section of the tumor specimen, using the same experimental setup described in **Figure 33**.

First, a single 14 μm thin section of this tumor specimen was extracted and placed on a glass slide and visualized on the microscope. The UCNP emission was obtained under a continuous 980 nm excitation (45 W/cm^2) and the tissue texture was captured using tissue autofluorescence under a 450 nm excitation light. A composite overlay of the high-resolution images obtained on the microscope is shown in **Figure *35***(a), illustrating two distinct sites with a higher concentration of UCNPs. The slide is then directly placed on the micro-chip, similar to the experiment depicted in **Figure 33**. The image captured by the micro-chip is shown in **Figure *35***(b) and is obtained using a pulsed 45 W/cm^2 980 nm excitation light. The two areas of UCNP are highlighted in **Figure *35***(a) and **Figure *35***(b), correspond to the regions identified on the microscope image. Each of the emission site in zones A and B is approximately 150 μm in diameter, and contains about 200 cells and the imaging sensor was able to resolve both sites with a signal-to-background ratio of 4.3 (or 21/4.9). The reduced SBR is explained by the reduced signal intensity in the 14 μm section of the specimen compared to the 2 mm thick unsectioned sample.

6 Conclusion

Intraoperative detection of microscopic residual disease has remained a persistent challenge for cancer resection surgeries, despite the fact that it significantly increases the rate of cancer recurrence if left behind. Miniaturization of large fluorescence microscopes and intraoperative imagers has remained the main obstacle in achieving fully guided surgery with a sensor that would be compatible with today's minimally invasive surgical procedures and complex tumor resection cavities. Limited by the large and rigid optical elements and equipment they require in order to resolve the image, fluorescence intraoperative imagers have failed to achieve millimeter scale miniaturization and still retain a relatively large form factor making them incompatible for universal and intraoperative resection imaging.

In this work we presented a new platform for intraoperative imaging, dispensing with conventional bulky optical devices in favor of completely optics-free silicon-based contact imager arrays. This is made possible by a synergistic combination of UCNPs with unique optical properties -long decay lifetimes, upconversion capability and no photobleaching- and custom integrated CMOS circuit design with a time-resolved array and a novel in-pixel background subtraction scheme. Using a time-resolved image acquisition scheme, we are able to capture the image without any optical filters, and as a result, scale down the imager size and thickness to less than 100 μm, if needed. We have leveraged the long emission lifetime of UCNPs and their immunity from photobleaching and long excitation wavelength to virtually eliminate tissue autofluorescence, greatly enhancing the signal-to-background ratio during the process. This ultra-thin imaging "skin" has also an embedded novel background subtraction scheme to mitigate the effects of the electrical background generated by the pulsed NIR (980 nm) excitation light, which creates interference on the CMOS photodiodes in the pixels. This subtraction scheme has been made possible by implementing a secondary photodiode to measure this local background and later adjust the main photosensor signal.

Table 1 includes a comparison of time-resolved biomedical sensors. While other works, as seen in Table 1, have also been able to achieve filter- and lens-less acquisition using beam orienting and long (>1.5 μm) NIR excitation wavelength [33] or correlated samplings of single-photon avalanche diodes [35], the performance of these platforms will significantly degrade when directly exposed to or excited with NIR-I light such as 980 nm, due to the penetration depth of NIR-I and the generation of long-lasting (millisecond range) background that will compromise signal integrity. As a result, they are incompatible with NIR-I imaging. Our novel dual-photosensor design architecture, however, ensures a relaxed limit on the orientation of the excitation source, and additionally, the background cancellation is no longer wavelength-specific and can be easily modified to be compatible with any given NIR wavelengths ($\lambda < 2$ μm), enabling optics-free imaging of targeted microscopic residual disease sites with upconverting nanoparticles in clinical settings.

The proposed imaging platform demonstrated being able to resolve targets with a clearance of 71 μm which represents a resolution of smaller than 10 cells, achieving enough spatial resolution for cancer staging purposes which only require margin assessment with 200 cells for microscopic residual disease detection[36].

The sensor was also able to resolve regions of UCNPs on a 2 mm thick tumor specimen directly placed on the micro-chip with no optics or focusing lenses -directly illuminated with an NIR-I (980 nm) laser- and achieved a signal-to-background ratio of 8, and was able to reduce the background by more than 20x, highlighting the critical role of the novel background adjustment

scheme. This performance was also maintained when imaging a single 14 μm section of the described specimen, and the background correction scheme resulted in an SBR of 4.3, with the reduction in SBR due to the smaller amounts of UCNPs per area and therefore a lower signal intensity.

This imager array achieved real-time frame rates (105 fps) and an output (measured) noise level of 0.36% full scale rms -where full scale is the count limit of the pixel, i.e., 600 a.u. The majority of this noise is contributed by the pixel circuitry but can be further attenuated by temporal averaging (averaging consecutive and repeated images), made possible thanks to the fact that UCNPs do not suffer from photobleaching. A fill factor of 47% was also achieved, with losses mainly attributed to in-pixel readout circuitry as well as the additional secondary photosensor.

Table 1. Comparison table of time-resolved biomedical sensors.

	IEEE TED'12 [37]	Theranostics'18 [33]	JSSC'12 [38]	JSSC'19 [35]	**This work**
Pixel array	256 × 256	36 × 80	32 × 32	192 × 128	36 × 80
Frame rate	15 fps	137 fps	20 fps	18.6 kfps	105 fps
Array size (mm × mm)	-	4.9 × 2.5	4 x 4	3.2 × 2.4	2.3 × 4.8
Fill factor	4.6%	64%	37%	13%	47%
Exc. wavelength	440 nm	1550 nm (NIR-II)	~610 nm	685 nm	any $\lambda < 2$ μm
Optics used	Filter and lens	None	Filter and lens	None	None
Noise (rms full scale)	-	0.4%	4.6%	-	0.36%

In summary, we demonstrated a chip-scale optics-free (filter- and lens-less) intraoperative imaging platform that utilizes the optical properties of lanthanide-based upconverting nanoparticles as molecular probes and leverages their uniquely long emission decay times and long excitation wavelength (980 nm) to perform time-resolved imaging in an environment where no tissue autofluorescence or photobleaching is present, opening the door to highly sensitive intraoperative imaging with a minimal form factor chip-scale imager.

7 Further Improvements

Further improvements of this platform may include the use of external planar micro-gratings to enhance image sharpness and optimization of the secondary photodiode size and shape to maximize emission signal captured on the main photodiode while still providing a reliable measurement of the local background.

The use of newer fabrication technologies such as SOI that allows isolation between substrate and bulk with the active region can also greatly benefit this design. Using SOI fabrication technologies can help relieve the system from the interference of the optically-induced charges in the bulk as they no longer can travel to the active region and no longer create interference on the photodiodes.

Using a fabrication technology with a higher level of doping for the substrate is also another alternative method in reducing the magnitude and lifetime of the substrate carriers during excitation.

8 Appendix: Use of Deep learning in 3D Reconstruction of Biological Specimens

In this section, we will discuss the implementation of a convolutional neural network (CNN) for spatial reconstruction of three-dimensional specimens using a single or two instances of the imager described in this work. This work has been conducted in cooperation with Rozhan Rabbani (UC Berkeley) and Asmay Gharia (UC Berkeley).

8.1 Deep learning and biomedical imaging

Visualizing the 3D location of fluorescently labeled tumor cells *in vivo* is critical for intraoperative navigation to identify tumors beneath the tissue surface and harbored in deeper sites such as lymph nodes. Conventionally, this process is done post-operation, in a laboratory setting by using targeted fluorescent probes and markers to identify any disease from the tumor bed sample. This process is extremely time consuming, and requires several days to return results, potentially putting the outcome of the treatment at risk. Recent engineering advances have resulted in several novel imaging platforms that allow this process intraoperatively and concurrent with the surgery [39]–[41]. However, these instruments are significantly cumbersome and not practical with today's minimally invasive surgical procedures, especially in complex and hard-to-access tumor cavities. While these large instruments rely on sizeable optics and lenses for their high resolution and reliability, they can't be miniaturized and made practical for surgical settings as a result of those inevitably rigid optical equipment. In addition, such large instruments naturally do not allow for assessment of treatment response, where cell migration into tissues is important to be monitored in real-time and outside surgical settings– another critical application. Therefore, a much less complex imaging platform with a smaller form factor is preferred. Recent advances in microscopy such as light-field microscopy [42]–[46] have enabled imaging of smaller features and imaging within the tissue. However, these methods require specialized optical equipment incompatible with a minimally invasive procedure.

Miniaturization of these platforms into electronic micro-imagers such as [23] enables placement of imagers in hard-to-access regions, unlocking the ability not only to visualize microscopic disease intraoperatively in cavities up to several millimeters deep, but also to monitor cell dynamics and assess treatment in real-time and in vivo, with a network of wirelessly powered implants. **Figure *36***(a) and **Figure *36***(b) illustrate how micro-imagers can provide a comprehensive visualization of the tumor of interest, with no disruption to the flow of operation or treatment for intraoperative and implantable applications such as in [29], [47], respectively. For intraoperative imaging, by rotating a surgical fiducial shown in **Figure *36***(a) multiple image acquisitions from different angles of view can be obtained. Similarly, a network of implantable imagers can capture images from different angles of the target as shown in **Figure *36***(b).

While these ultra-small imagers are easy to integrate into surgical environments, unlike their larger counterparts, they lack a high enough resolution. Reducing the form-factor of the imager imposes a stringent limit on the size of the optical filters and focusing lenses being used, limiting their performance and the image resolution. Smaller imagers needed for in vivo use have often a higher background level (due to the lower rejection performance of smaller optical filters) and a lower focusing capability (due to the limited numerical aperture of their lenses). Safety limits also restrict the total photon budget allowed within the system, further constraining the capabilities of fluorescence microscopy in vivo. To obtain reliable 3D information using imaging instruments with these small form-factors, enhanced custom-made optical filters and lenses are required to replicate the same performance as their larger counterparts. These, however, are often difficult to manufacture, and image quality remains suboptimal relative to bench-top microscopes. Therefore, computational techniques that can enhance images from small form factor devices are in need.

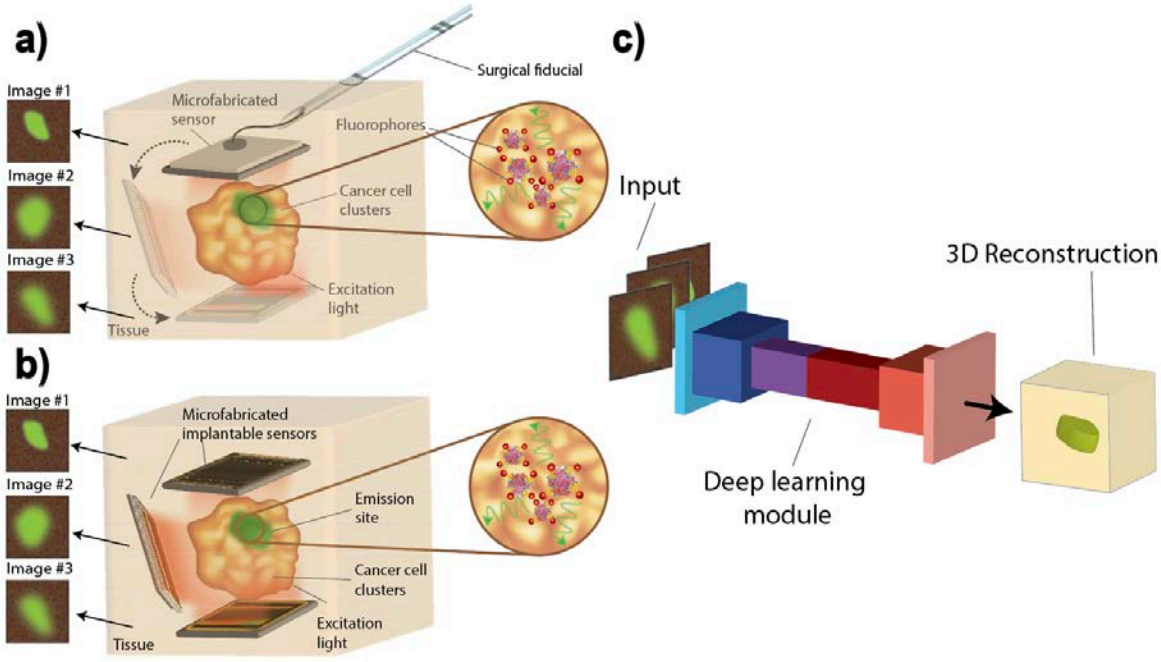


Figure 36. Concept of multiple visualization of the tumor using micro-imagers: (**a**) multiple images taken by rotating the micro-imager intraoperatively. (**b**) network of implantable micro-imagers to capture multiple fields of view. (c) combination of neural networks and micro-imagers enabling 3D visualization and resolution enhancement.

Conventional image processing techniques involve variations of deconvolution, surface projection algorithms and noise enhancement methods, and the core of all these methods relies on a linear transformation of the image, that does not depend on anything other than the raw image data and the point spread function (PSF) of the imaging device and is procedurally blind to any prior knowledge of the specimen being observed.

The PSF, also known as the transfer function of the imaging system in spatial domain, describes the response of an imaging system to a point source of illumination [48]. In any linear image formation process such as fluorescence microscopy, the final image is a linear superposition of a series of point sources convolved with the PSF. Therefore, the original object can be retrieved by deconvolving the image with PSF of the imager.

One of the drawbacks of the deconvolution method is the calibration required for deriving parameters of the deconvolution function for images taken from each depth which limits image processing speed. This method poses difficulties for recovering images consisting of overlaid cell

foci from different depths since every parameter requires optimization for a certain imaging depth. Moreover, the PSF, as a low pass transfer function, removes high frequency components of the original image, hence results in loss of sharpness, hindering generation of a fully recovered image after deconvolution. Even for cases where the image is mainly attenuated by the PSF response and still retains most of the high frequency components, applying inverse PSF amplifies high frequency noise degrading recovery of the original image. Much like deconvolution, every other linear image processing technique will suffer from similar issues.

To circumvent these limitations, a more agile approach is needed, namely a non-linear post-acquisition processing module that could incorporate the physiological and spatial information of similar tissue specimens within itself. The module can leverage this additional knowledge to restore the sharpness and resolution to the micro-imager's suboptimal images, and provide insight into the 3D position of cells. Of all available architectures, deep learning modules can be by far the most adept at capturing physiological knowledge of cellular images.

Deep learning combines multiple layers of non-linear transformations, superposed with a complex yet structured network of coefficients to create powerful processing modules that can perform highly complex tasks, such as image enhancement, image classification and feature extraction. Deep learning allows breaking the tradeoffs of fluorescence microscopy and using the computational models to augment hardware complexity and improve upon optical limits, by using a large collection of training data to build the network [49]. Using adaptive network architectures such as residual neural networks (ResNets[50]) and convolutional neural networks (CNNs[51]), we have included here several applications for cancer imaging utilizing deep learning to enhance the resolution and capability of custom-made micro-imagers [33], [52]–[55]. **Figure *36***(c) demonstrates how the symbiotic combination of neural networks and micro-imagers can not only restore the sharpness and resolution of the image, but also create a 3D visualization of the specimen at no additional hardware cost and create a highly reliable intraoperative imaging platform for margin assessment.

A key challenge to using a neural network in 3D cellular imaging of tissue is compiling a training dataset, since training deep neural networks requires access to a large set of training data from diseased tissue for each specific application. The impracticality of obtaining a large dataset, taken at varying depths, from tissue motivates synthesis of a diverse dataset of tumor cell images based on the morphology of real-life tissue samples, to leverage prior knowledge of the tumor cells. The synthesis method needs to be parameterizable to allow generation of an arbitrary large dataset by random selections of parameters that lead to images which are accurate representations of actual cells.

To address this, we first discuss a method to generate a large training set mimicking real-life specimens (a single layer of cells on slide), thus allowing the deep neural networks to be trained, as demonstrated in [56]. To replicate the 3D structure of the tumor, stacks of multiple layers of cells with 250 μm spacing within 1 mm from the sensor are generated. Since the lensless custom imager in [22], [24] is designed for contact imaging of tumor margins, 1 mm was set as the limit to demonstrate proof of concept. Next, we demonstrate 3 modules, tasked with identifying the depth of cell clusters -measured as the distance of the sample from the imager-, deblurring and enhancing the sharpness of the image, and finally, detecting cell presence within each layer of the specimen in 3D stacks. The last module incorporates a novel method of imaging using not one but two sensors viewing tissue from different angles to allow for three-dimensional imaging of the sample and providing insights into the spatial distribution of the cells in the sample.

8.2 Dataset synthesis

To generate images identical to real-life cell foci, a coherent gradient noise generation method referred to as Perlin noise is used [57]. Perlin noise is a technique generally used to create natural appearing procedural textures, such as marble, wood, cloud textures for motion picture visual effects [58].

To locate the tumor cell foci, a binary matrix is generated to represent a tumor mask with high values (>0.5) indicative of tumor and low values (<0.5) for non-cancerous background. Exploiting the natural structure of Perlin noise, a smooth cellular location map is achieved, populated with signal and background intensity values. Once the location of the cell foci is determined with the tumor mask, a tumor image is rendered by producing in-pixel signal and background intensity values based on mean and variance of real tumor images. A close correlation with real images is ascertained by making sure the statistical parameters stem from a representative range of parameters resulting from real data SNR calculations demonstrated in [56].

Size of the tumor image is chosen to be 51x51, a form factor that parallels the lensless chip-scale CMOS imager for in vivo intraoperative imaging of cancer[23].

8.3 Modules implemented

8.3.1 Depth estimator module

Nonlinear networks for depth estimation are often used in non-medical contexts, for traffic, navigation and security purposes [59]–[62], and have proven to be quite capable of identifying spatial features and depths if properly trained prior to testing. To this end, we have built a convolutional neural network combined with a fully connected (FC) network to estimate the depth of images that emulate the custom-made imager's resolution and PSF at various depths, randomly generated from 50 µm to 1.95 mm, with 100 µm increments.

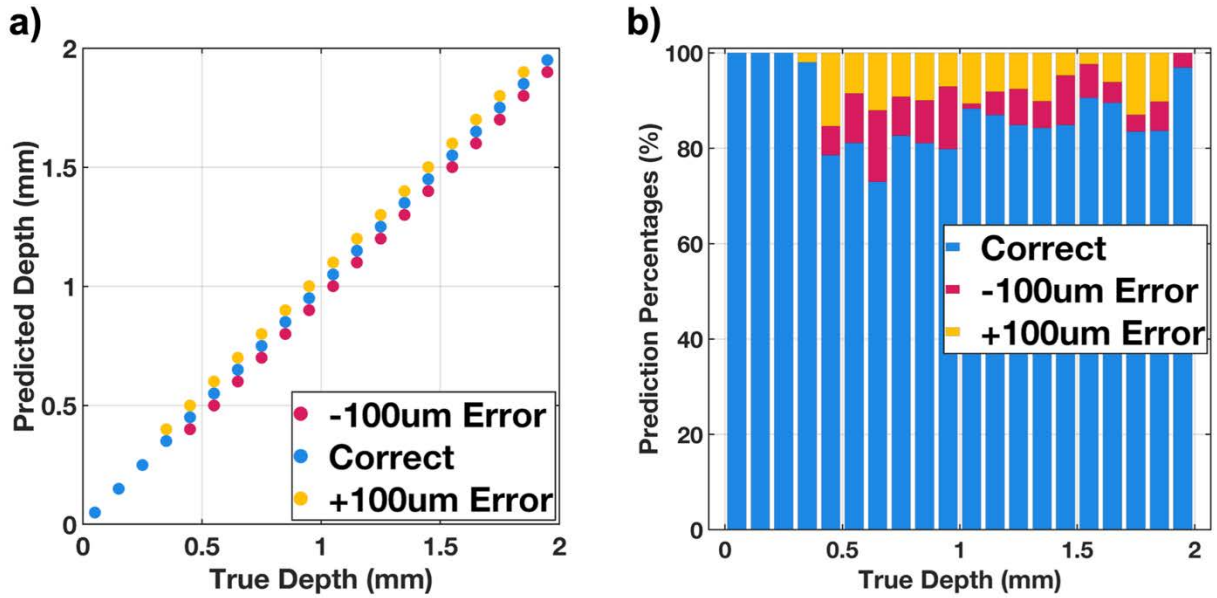


Figure 37. Performance of the depth estimator module: (a) predicted values for every depth in the test dataset. (b) test accuracy and loss for depths ranging from 50 µm to 1.95 mm with 100 µm step sizes.

The training dataset for this network has been compiled by initially generating raw and un-blurred images of a single layer of cells, after which the PSF of the microchip imager at different

depths was applied onto them. The depths were randomly sampled from the range (0, 2 mm), with 100 μm step-size, thus allowing for 20 distinct depths.

After training, the module was tested on 1000 randomly sampled images at various depths ranging from 50 μm to 1.95 mm, and the prediction performance is shown in **Figure *37***. **Figure *37***(a) shows the occurrence of all the predictions in the test dataset, and **Figure *37***(b) illustrates the detailed statistical performance of it. The module was able to identify the depth of the test images with a maximum error of 100 μm, enough to enable resection of tumors, even below the tissue surface.

If the depth resolution is assumed to be 100 μm, then the accuracy rate of this module would be about 87%, with the accuracy growing closer to 100% as the specimen is closer and closer to the imager.

8.3.2 Deblurring module

Obtaining high resolution images while maintaining a small size for the imager requires information beyond the mere raw data of the image, and an advanced knowledge of image contents. Knowing and having prior information about the underlying image data allows a more efficient recovery of the un-blurred image and insightful clinical information. A convolution/deconvolution neural network can absorb this information and embed it into the network using an appropriate training dataset.

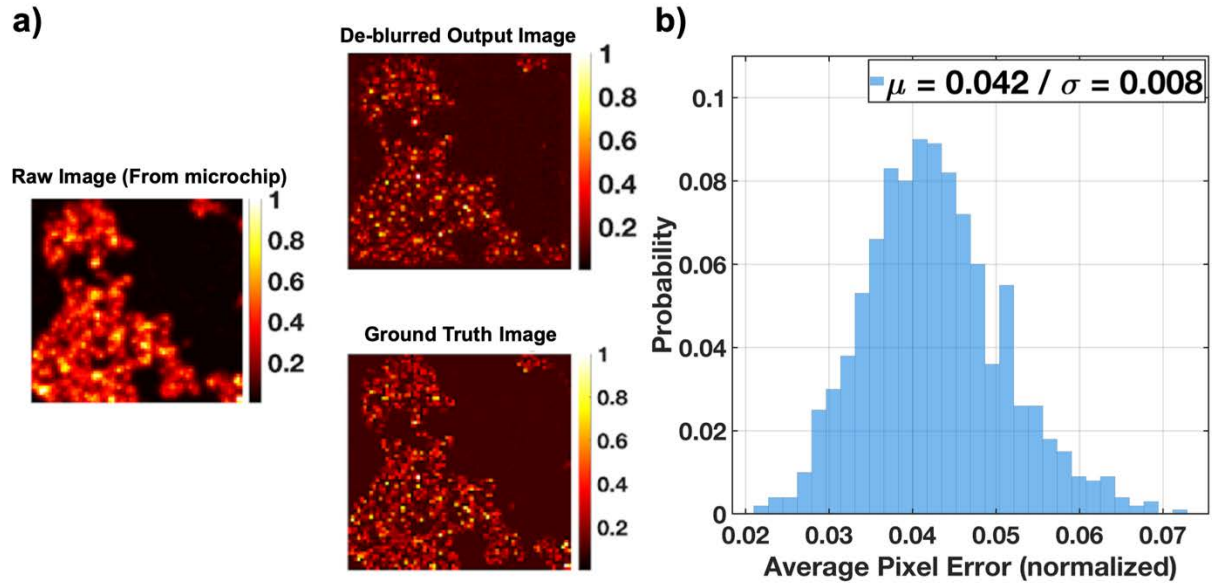


Figure 38. Performance of the deblurring module: (a) network input and output images for a test sample in comparison with the corresponding ground truth image. (b) distribution of average pixel error for 1000 test samples with a mean error of 4.2%.

Here we created a 6-layer CNN trained to enhance and deblur corrupted images obtained with the custom-made imager. The training dataset used for this module was a compilation of single layers of emulated cell images to which the PSF of the custom-made imager was subsequently applied at different and randomly generated depths, ranging from 0 to 1 mm.

We trained a 6-layer (3 convolution/3 deconvolution) network on a training dataset and tested the performance of the module on 1000 test samples and **Figure *38*** illustrates the

performance of the module. This module can extract the un-blurred image from a blurred input and restore sharpness and original resolution to the image, by relying on embedded prior image information and content-aware non-linear transformations. A sample test performance is shown in **Figure *38***(a).

The statistical accuracy of this module is illustrated in **Figure *38***(b) which shows the average normalized deviation of pixels from their correct value is about 0.04, which translates to 4% estimation error.

8.3.3 Cell detector module

Building on the module for depth estimation described in (8.3.1), this module extends to localization of cells from different layers of the tissue and identification of their corresponding depth. The underlying advantage of multi-layer depth estimation exploiting machine learning models is distinguishing dim clusters of cells closer to the surface of the imager from bright cellular response far away from the device, a task hard to achieve without further processing. Multi-layer cell detection from a single 2D image with a precision better than 500 µm provides surgeons with high enough resolution to investigate the tumor bed.

Extracting multi-level depth information from the planar images of our lensless microscope on-chip eliminates the need for bulky optical lenses [63]–[65]. We have developed separate modules for detecting nonoverlapping and overlapping multilayer clusters of cancer cells. To generate the training dataset for nonoverlapping stacks of cells, pairs of emulated cell images from 2 different depth values are chosen randomly with a minimum difference of 500 µm from 0 to 1 mm. Initial cell images from each layer are convolved with the PSF of the custom-made imager and added together spatially to form a multilayer image. The overlapping regions are subtracted from the stacked image to ensure separation of the cells from each layer.

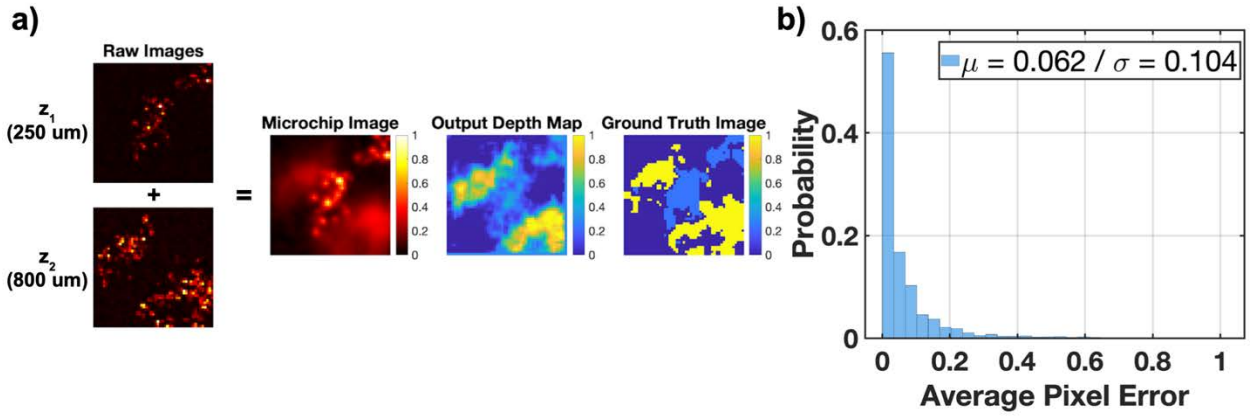


Figure 39. Performance of the cell detector module for non-overlapping stacks of cells: (a) Raw images at each depth before applying the PSF, network input and output images for a test sample and the corresponding ground truth image. (b) Distribution of average pixel error for test samples with a mean of 6.2% for 1000 test samples.

To extract the location of cell foci in the two layers of cells, a 6-layer (3 convolution/3 deconvolution) network is trained on the dataset and the performance is evaluated on 1000 test samples. **Figure *39***(a) shows the overlaid image from the stacks, the corresponding depth map and finally the depth map predicted by the model. **Figure *39***(b) depicts the error distribution of the depth map predicted by this network with an average normalized deviation of pixels from their

correct value to be about 6% with a standard deviation of 10%. Having established a model that successfully identifies two-layer depth map of tumor cells within a normal tissue background, we aim to show that we can generalize 3D localization to a more comprehensive case discussed in the next section.

In addition to detection of cells in multilayer segregated clusters, 3D information from a more complicated structure with overlapping stacks of cells is explored in this work. However, the accuracy of recovering 3D information using a single micro-imager is going to be very limited when extracting spatial information from farther layers. The decreasing optical signal and increased scattering and absorption and more importantly the PSF non-ideality will result in a very low signal-to-background ratio in the farther layers.

As a result, and to mitigate the high error rate of the model, we have proposed adding a second sensor to the imaging system, which is made possible by the ultra-small form factor of the sensor itself, allowing it to become fully implantable (if required) and therefore surgically practical. In our emulated experiments, the two sensors are placed 1 mm apart, on both sides of the emulated three-dimensional tissue.

Both the single and dual-sensor modules consist of a 6-layer CNN (3 convolution and 3 deconvolution layers) and their outputs are 4 binary input-sized layers depicting cell presence in each one, where [>0.5] indicates cell presence and [<0.5] indicates absence of cells. Due to the binary nature of the outputs, we can evaluate the accuracy of the outputs in terms of pixels incorrectly “labeled” (“existence” or “absence” of cells)- a metric we will later use to compare their performances.

For a reliable 3D imaging of tissue, it is necessary that the module maintains its sensitivity and specificity performance across the entire depth of the specimen, and as such, the limited performance of this module will preclude it from being used to acquire reliable 3D information and perform deep tissue imaging on samples that are more than a few hundred microns thick. In this work, sensitivity is defined by the ratio of pixels accurately predicting presence of cells over the total number of cancer cells in the ground truth images. Specificity, on the other hand, evaluates performance of the model in predicting absence of cells in the pixels indicated by the ground truth images to be empty of cells.

Amongst various applications of cellular level depth estimation, one of the more critical applications of microscopic imagers in oncology is to monitor and observe the movements and dynamics of the cells which represent the real-time response of tissues to therapy. The speed, direction, and features of the clusters of cells experiencing those dynamics are of significant clinical value, yet due to the complexity, cumbersomeness and optical limitations of intraoperative imagers, clinicians often have no choice but to do away with them. We here demonstrate in our final module a 3D-reconstruction model architecture that is capable of capturing these dynamics. The model is used to evaluate the sensitivity to cell dynamics and movement across layers and verified quantitatively with test samples.

CNN with single sensor: The first multi-layer cell detection module is based on a single sensor, observing a stack of 4 layers of cells that are randomly space between 0 and 1 mm away from the sensor itself, where each layer is at least 200 μm away from the adjacent layer- to allow full coverage of the [0,1 mm] three-dimensional space with only 4 layers. Individual layers are uncorrelated, and have randomly generated intensities and background levels, allowing for a realistic emulation of tissue. Upon applying the corresponding PSF to each layer, the 4 images are summed and combined into one final image that constitutes the sensor's output and will serve as raw input to the module.

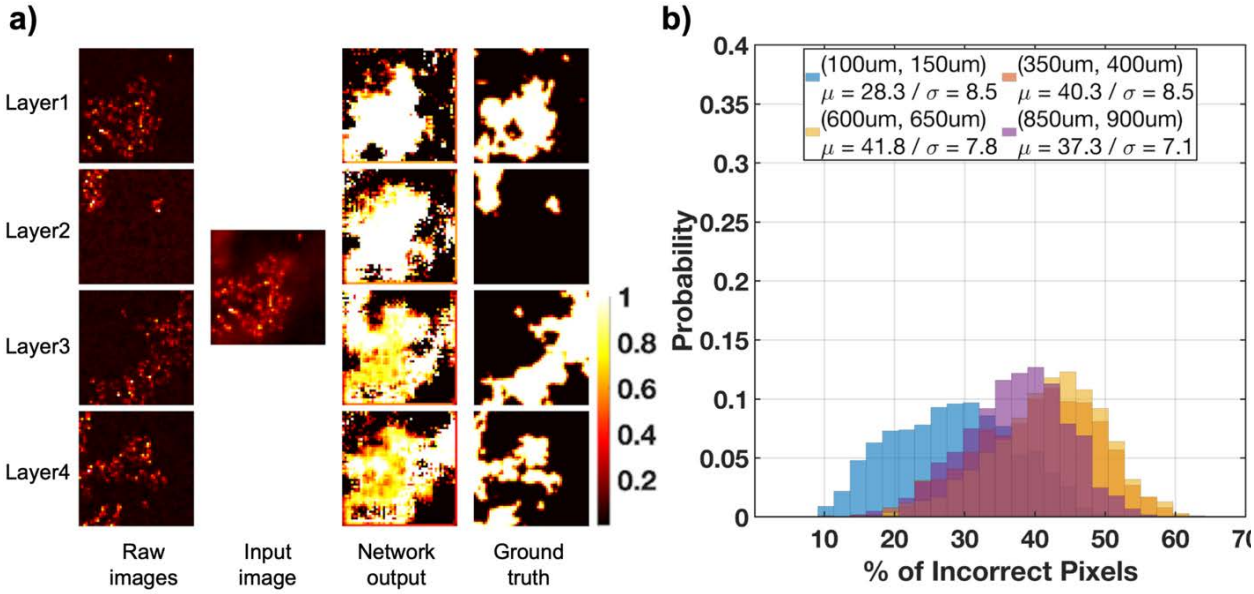


Figure 40. Test results of the CNN model with a single sensor: (**a**) Overlaid input image from 4 raw images corresponding to each layer, network output images and ground truth depth maps for each layer. (**b**) Distribution of average pixel error for test samples for each layer (starting from the closest one to the sensor) with averaged error rates of 28.3%, 40.3%, 41.8% and 37.3% for layers 1 to 4 respectively.

After training, the module evaluated over 1000 distinct test inputs, and the corresponding input and outputs for one sample image are shown in **Figure 40**. The distribution of the performance is also shown in **Figure 40**. Our performance metric reveals that the first layer, that is the closest to the sensor, has a lower error rate of 28%, and the performance degrades with farther layers, with the rate of inaccuracy remaining well above 37%.

The distribution of the performance illustrated in **Figure 40**(b) is shaped by random overlap of cells in the four layers. A deeper analysis of the individual data confirmed that the cases with lower counts of incorrect pixels observed in the far layers are caused by the subsequent cell layers overlapping significantly with the closest one, resulting in special cases and lower than usual error rates.

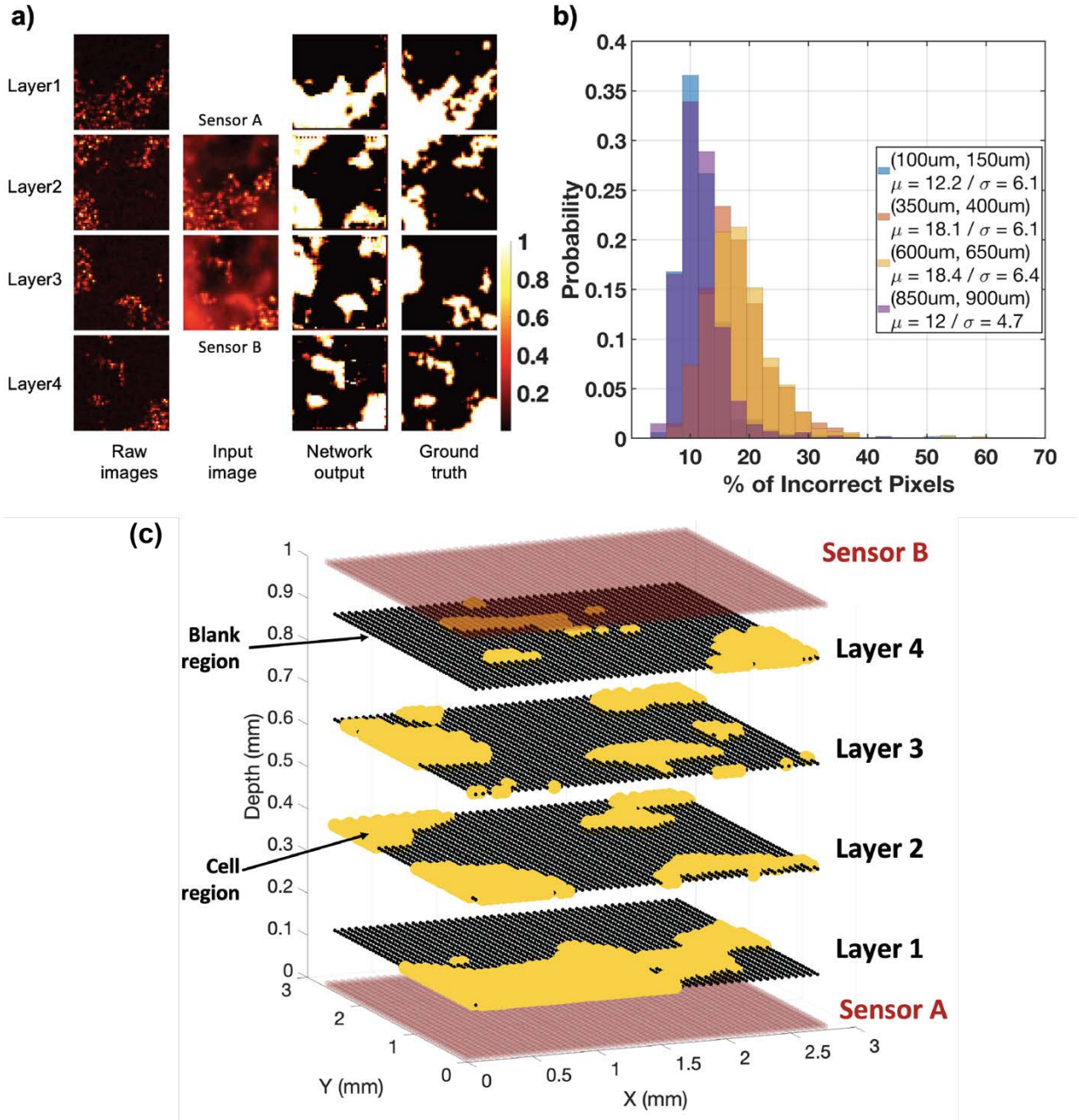


Figure 41. Evaluation of the CNN with two sensors: (**a**) overlaid input images from 4 raw images corresponding to each layer before applying the PSF, network output images and ground truth depth maps. (**b**) distribution of average pixel error for each layer with averaged error rates of 12.2%, 18.1%, 18.4% and 12% for layers 1 to 4, respectively. **(c)** spatial (3D) reconstruction (using network outputs) of the sample test input, where black, yellow and red respectively represent blank (empty of cells) spaces, regions containing cells, and sensor locations.

CNN with two sensors: The effect of adding the second imager, which is possible with the current form factors of these imagers, to the opposite side of the target under test is explored with the generalized 4-layer case to quantify the improvement of the accuracy.

Similar to its single-sensor counterpart, and using the exact same network, we evaluated the module over 1000 distinct test inputs and the corresponding inputs and outputs for one sample

image set is shown in **Figure *41***(a). The impact of adding a second sensor can be observed by comparing the outputs shown in **Figure 40**(a) and **Figure *41***(a), quantified in **Figure *41***(b). **Figure *41***(b) illustrates the performance distribution of the 2-sensor module, and as expected, the first and last layers have very similar performances, as do the two middle ones, and this network can achieve a 12% error rate (in the two closest layers), which is less than half of the error rate of its single-sensor counterpart. Adding a second sensor reduces the error in the two middle layers – farthest from the sensor – reducing it from 40% to 18%.

Upon extracting and identifying the regions in each layer that include cancer cells, we are then able to reconstruct a spatial representation of the stacked sample in space and provide a 3D visualization of the specimen being imaged. Illustrated in **Figure *41***(c), using the outputs shown in **Figure *41***(a), we have reconstructed the stacked sample, identifying the zones where cancer cells were detected in each layer. For a complete representation of the proposed imaging platform, the two sensors are also shown in **Figure *41***(c), separated by the 1 mm thick stack of cell layers in between.

ResNet+CNN with two sensors: Here we present a module that can not only identify cell clusters across all depths but is also very sensitive to small changes in the specimen, i.e., resulting from cluster movements. Using a more enhanced neural network, this module is able to identify cluster movements between layers and help visualize the dynamics within the specimen. In the previous section we introduced a two-sensor architecture to significantly improve the accuracy of the module. Here we preceded a 2-layer CNN network with a pre-trained 18-layer ResNet architecture[66], and after training, evaluated the performance of the compounded network on the test dataset. The architecture of the network and its input images and output depth maps are shown in **Figure 42**.

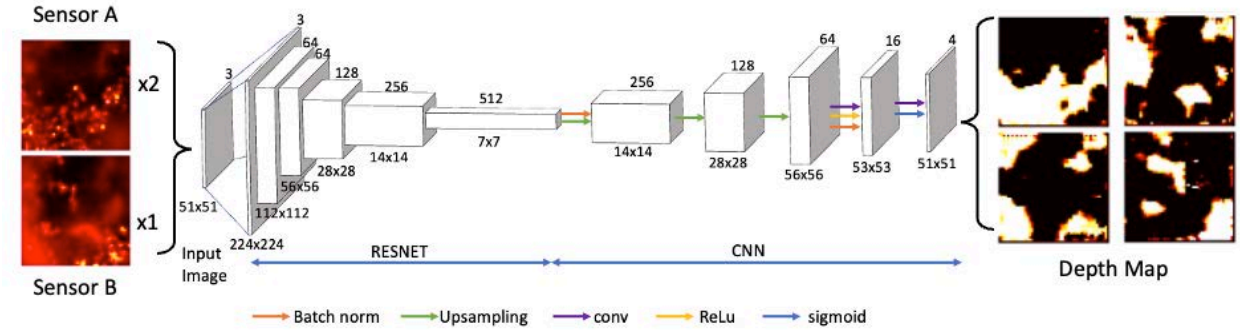


Figure 42. Architecture of the deep neural network model consisting of 18-layer ResNet and CNN with the corresponding input images from 2 sensors and the output depth maps for 4 layers. A replica of the image from sensor A is added to the 2 input images to comply with the 3-channel input of ResNet.

The test dataset includes 100 distinct and randomly generated group images, in which a cluster of cells within the three-dimensional space moves across the different layers, going from sensor A towards sensor B, thus replicating a physiological dynamic of a real-life migrating cluster of cells (such as immune cells migrating into a tumor, or a metastatic deposit migrating or dividing within tissue). **Figure 43**(a) shows the diagram of the described scenario and how the cluster (and only the cluster) moves across the layers.

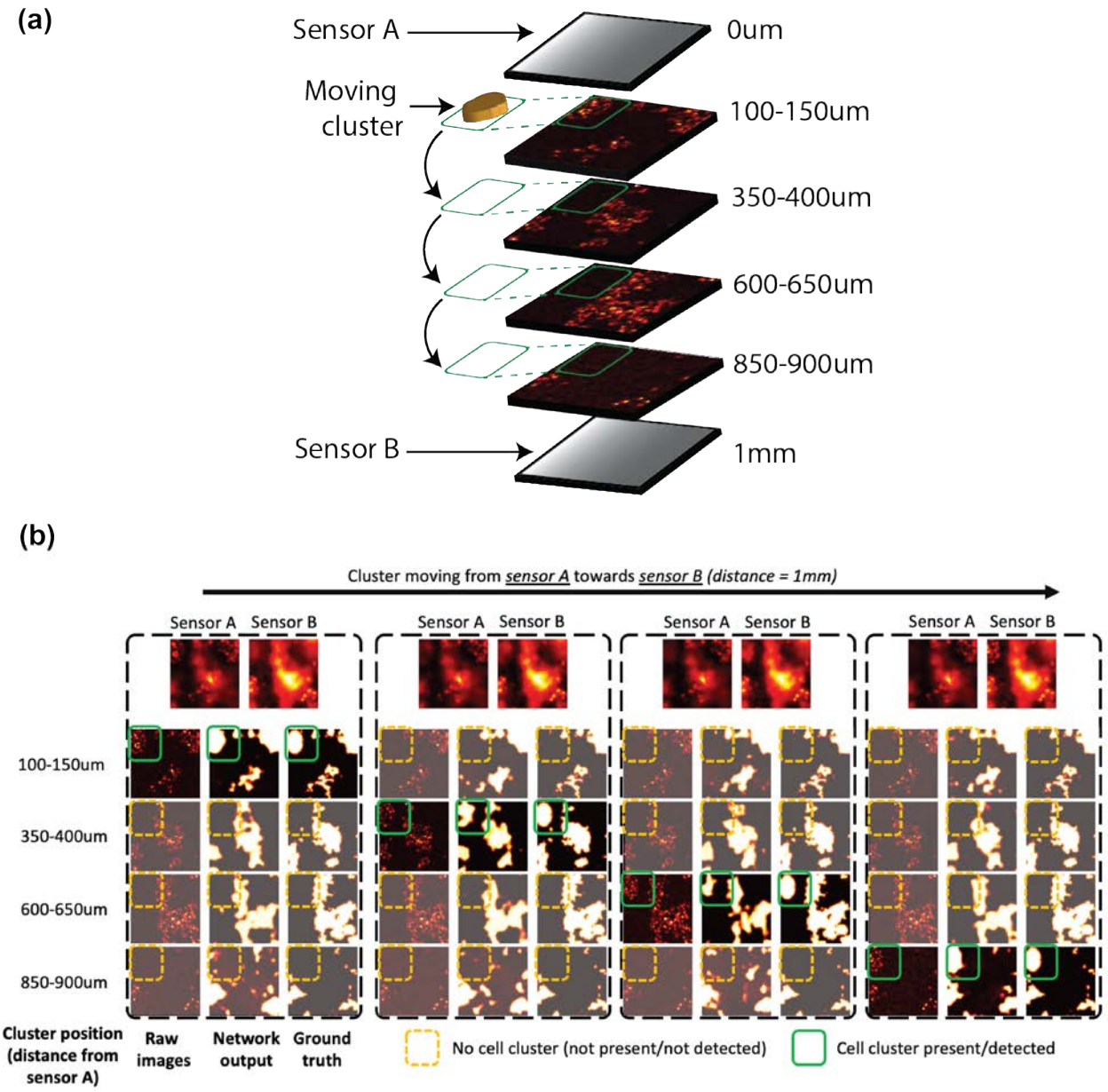


Figure 43. **(a)** test setup for modeling dynamics of a moving cell cluster. **(b)** outputs of the compounded module with the moving cell foci at each layer including the deblurred images at each depth, the two sensor images captured that serve as input to the module, the network outputs and the ground truth cell maps at each depth.

The performance distribution of this module is shown in **Figure 44** and it can be seen that the two closest layers have a very low error rate (~11%) while the two middle ones have slightly higher rates (~16%). However, both show noticeable improvements compared to the CNN-only network case.

The performance and outputs of the compounded module is shown in **Figure 43**(b). **Figure 43**b) includes 4 sections, each one of which includes the raw (unblurred) images at different depths, the network outputs, the ground truth image and the two sensor images captured that serve as input to the module. The cluster of cells marked in **Figure 43**(b) moves across the 4 layers (from left to

right), altering the 2 sensor images every time a layer change occurs. The module tracks the cluster with an average sensitivity of 72.6% and specificity of 91.7% for all of the 4 layers as shown in **Figure 45**(a). The Receiver Operating Characteristic (ROC) for average performance of the model across all layers is shown in **Figure 45**(b).

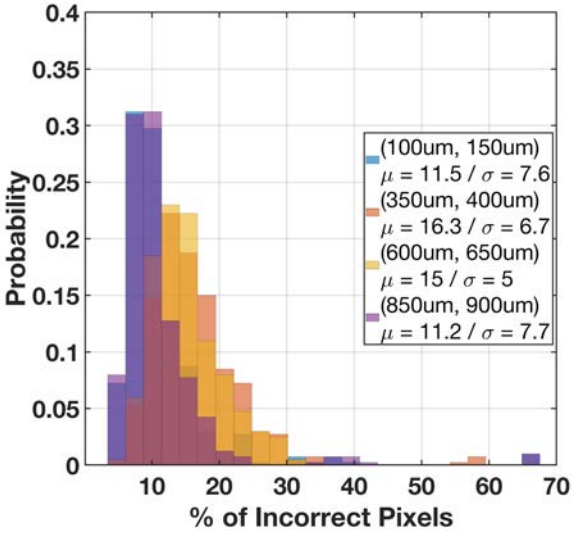


Figure 44. Performance of the ResNet+CNN model in identifying cell locations for each layer with average error rates of 11.5%, 16.3%, 15% and 11.2% for layers 1-4.

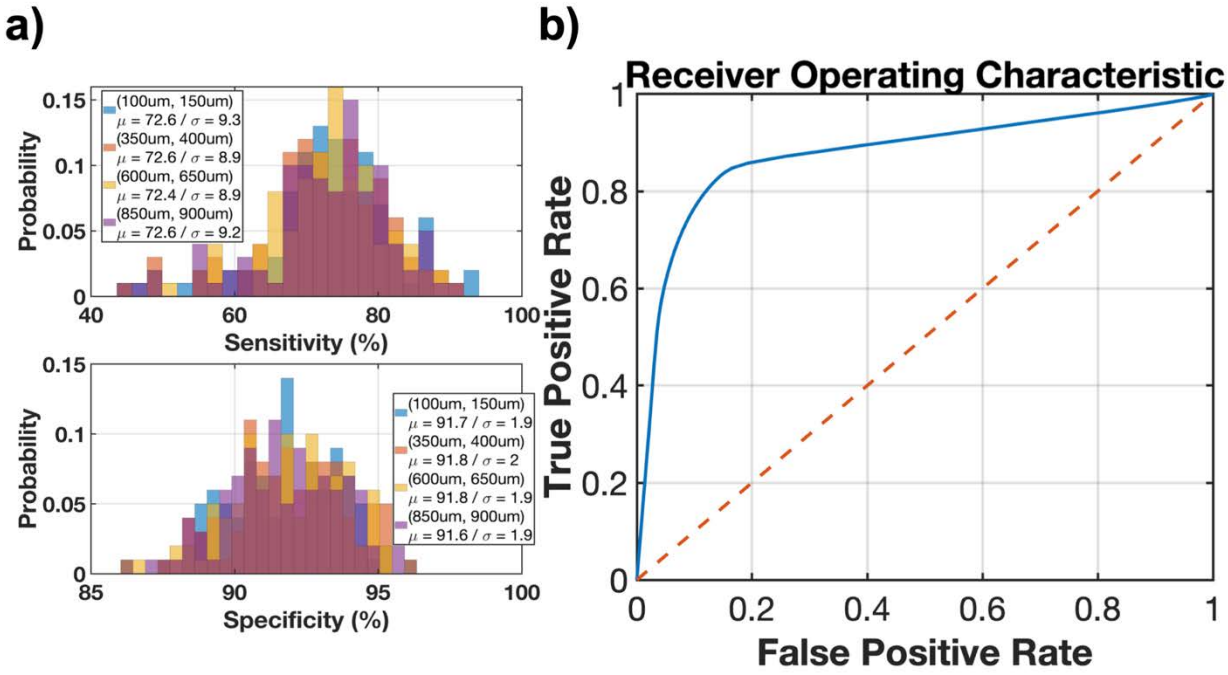


Figure 45. Performance of the compounded module with the moving cell foci: **(a)** sensitivity and specificity of the ResNet+CNN with 2-sensor network in detecting dynamics of the moving cell cluster for each layer. **(b)** receiver Operating Characteristic (ROC) of the model averaged for all

8.4 Comparison of networks

The modules presented in this work demonstrate the capabilities of neural networks in providing real-time enhanced image resolution and valuable 3D clinical information from currently available imaging systems and platforms. This project incorporates 3 modules featuring identification of the depth of a single image, deblurring and enhancement, and localizing cancer cell foci within each layer in both overlapping and non-overlapping stacks of cells. A novel method to synthesize a large representative dataset of real-life chip-scale images of the tumor (developed by Asmay Gharia - UC Berkeley) is utilized for training deep neural networks. We quantifiably demonstrated how placing two sensors on both sides of a tissue stack would improve the error rate of our depth estimation algorithm compared to the case with only sensor. The performance of the three modules described in this work have been summarized in Table 2, illustrating that the addition of a second sensor to a 6-layer CNN significantly reduces the error rate in the middle layers -by up to 55%. In addition to that, we also trained a 2-layer CNN preceded by an 18-layer pretrained ResNet and the introduction of the ResNet network further reduced the average error rate from 15.2% to 13.5%.

Table 2. Performance summary of 3 different trained depth estimation modules.

Architecture	6-layer CNN	6-layer CNN	18-layer ResNet + 2-layer CNN
Application	Depth estimation (3D reconstruction)		
No. of sensors	1	2	2
Layer 1 error rate (%)	28.3%	12.2%	11.5%
Layer 2 error rate (%)	40.3%	18.1%	16.3%
Layer 3 error rate (%)	41.8%	18.4%	15.0%
Layer 4 error rate (%)	37.3%	12.0%	11.2%
Avg. error rate	36.9%	15.2%	13.5%

Table 3 summarizes the performance of the image enhancement and depth estimation techniques presented in this work in comparison with state-of-the-art supervised monocular depth estimation models for single images from NYU depth dataset[67]. Depth estimation for indoor environments eliminates the need for detecting stacks of overlapping objects ([68]–[70]) while it is instrumental in detecting cell clusters in tissue samples. The cellular level depth estimation work presented in [71] relies on multiple acquisitions from different depths of focus, challenging for intraoperative applications. This work achieved accuracies of 87% for single depth estimation for depths ranging from 0 to 2 mm, 95.8% for deblurring and resolution enhancement of images taken up to 1 mm away from the imager, 93.8% and 86.5% for cell localization and depth estimation of non-overlapping and overlapping stacks of multiple layers of cells within 1 mm of the device,

respectively. To verify the performance of the model in dynamic applications, the network was also tested for identification of moving cell foci within multi-layer stacks of cells. The module was able to identify presence of the cluster with a sensitivity and specificity of 72.6% and 91.7%, respectively, achieving minimal "bleed through" between adjacent layers. Having demonstrated the synergetic performance of the system containing our customized image sensor[23] with the computation power of neural networks, for future work, a more comprehensive study is needed to optimize architecture of the neural network using structurers such as encoder-decoder based CNNs[72] or self-attention feedback networks[73].

Table 3. Comparison of this work with current depth estimation models.

	This work				**[35]**	**[36]**	**[37]**	**[38]**
Application	Single-layer depth estimation	Deblurring	Multi-layer depth estimation		Depth estimation	Depth estimation	Depth estimation	Cell invasion depth estimation
			Non-overlapping	Overlapping				
Architecture	CNN	CNN	CNN	RES + CNN	RES + CNN	Multi-scale deep network	Encoder-decoder	Gradient convergence
Single acquisition	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Dataset	Synthesized 2D images	Synthesized 2D images	Synthesized 2D images	Synthesized 2D images	NYU depth	NYU depth	NYU depth	3D cancer cell invasion assays
Average accuracy (%)	**87**	**95.8**	**93.8**	**86.5**	91.7	82.3	95.7	89.6

There are several limitations to our work. While we ensured our datasets replicated as close as possible to a real-life sample set, a synthesized dataset was used. However, our goal is to demonstrate proof of concept with this technique, and this approach can be repeated for any available cell imaging data set by retraining the neural networks on that dataset. Nonetheless, despite the synthetic dataset, the final model (ResNet+CNN) is still able to achieve a high level of performance when applied to real-life images, as seen in **Figure 46**. While gathering a large size of real-life images of cancer cells is beyond the scope of this work, we have applied the module on a limited number of cancer cell slides. These specimens have first been imaged on high resolution fluorescence microscope (shown in the first column in **Figure 46**), after which each one has been assigned a randomly selected depth and was applied the corresponding PSF. After applying the PSF, all 4 layers were combined (added) to generate the sensor image. A similar procedure is also carried out for the second sensor. The sensor images obtained are shown in the second column of **Figure 46**. The network output -illustrating regions where cells were detected in each layer- is shown in the third column, and an overlay composite image of the outputs with

the microscope images is also showing in the rightmost column in **Figure 46**, showing an almost perfect level of localized cell detection.

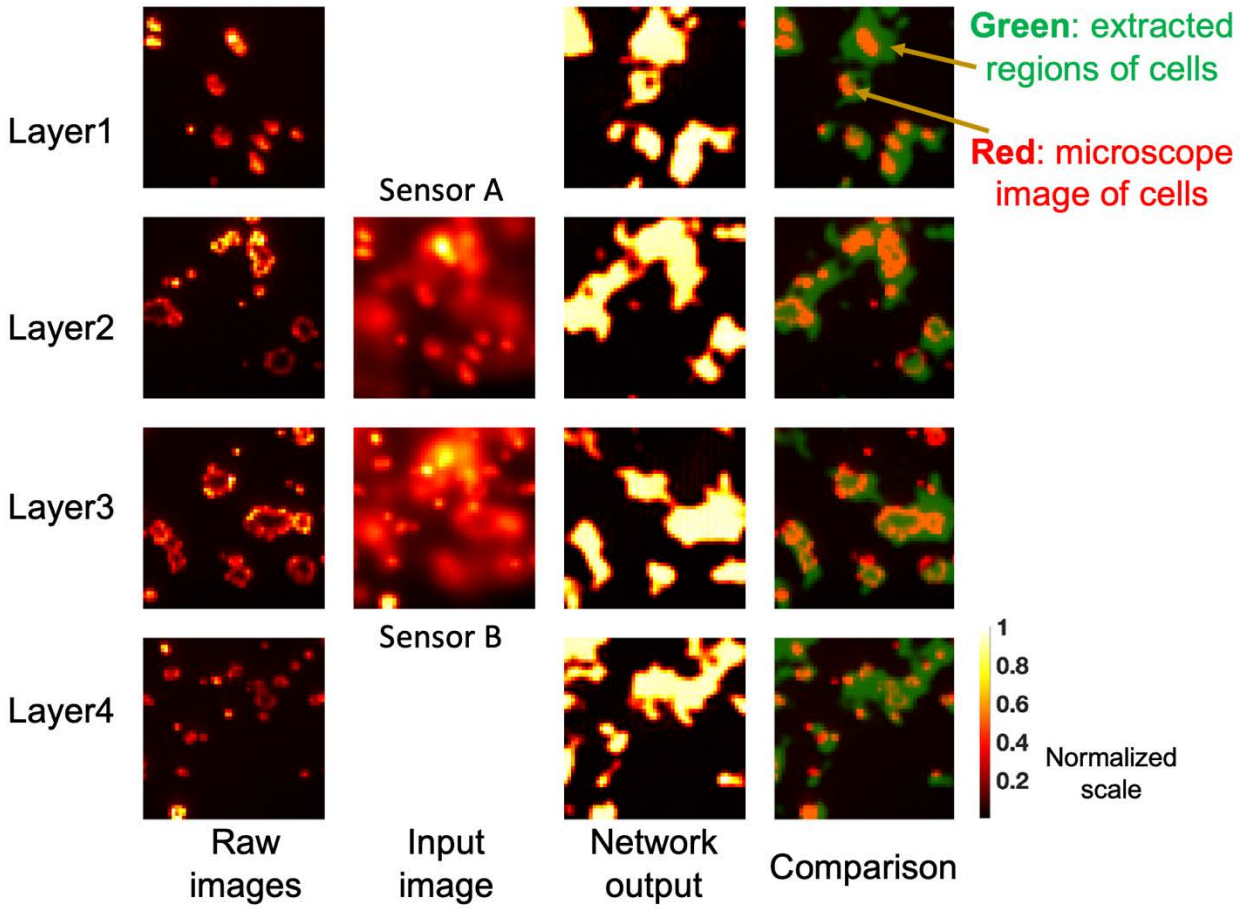


Figure 46. Evaluation of the CNN with two sensors on real-life cancer cell slides: (from leftmost to rightmost column) input images from 4 raw microscope images corresponding to each layer before applying the PSF, sensor images, network output images showing regions where cells were detected, and overlay composite image of network output and microscope images for each corresponding layer.

An additional limitation of this work involves a partial modeling of optical path non-idealities. While the PSF accounts for the optical divergence of the light path, it does not take account for the scattering and absorption of the tissue medium. A further enhanced version of the module can be trained on a dataset that incorporates the effect and impact of scattering, absorption, motion deformation, non-uniformity of the laser beam and additive noise in the tissue as described in [74]. While an improvement, those effects can however largely be mitigated by using a longer wavelength process with the corresponding fluorophores, such as an ICG or IR-800 dye, as reported in [75]–[77].

The superior representation capability of neural networks also allows for a far more complex network of sensors to be used than only two, and thus pushing the limits of detection and resolution even further than what has been presented in this work. Having readily available ultra-small sensors to be intraoperatively used, we can ultimately harness the power of machine learning in extracting valuable 3D information from a collection of easily obtained raw images.

Bibliography

[1] H. Bartelink *et al.*, "Recurrence Rates after Treatment of Breast Cancer with Standard Radiotherapy with or without Additional Radiation," *New England Journal of Medicine*, vol. 345, no. 19, pp. 1378–1387, Nov. 2001, doi: 10.1056/NEJMOA010874.

[2] S. E. Singletary, "Surgical margins in patients with early-stage breast cancer treated with breast conservation therapy," *The American Journal of Surgery*, vol. 184, no. 5, pp. 383–393, Nov. 2002, doi: 10.1016/S0002-9610(02)01012-7.

[3] A. N. Vis, F. H. Schröder, and T. H. van der Kwast, "The Actual Value of the Surgical Margin Status as a Predictor of Disease Progression in Men with Early Prostate Cancer," *European Urology*, vol. 50, no. 2, pp. 258–265, Aug. 2006, doi: 10.1016/J.EURURO.2005.11.030.

[4] O. RK *et al.*, "Positive Surgical Margins in the 10 Most Common Solid Cancers," *Scientific reports*, vol. 8, no. 1, Dec. 2018, doi: 10.1038/S41598-018-23403-5.

[5] T. A. Buchholz *et al.*, "Margins for breast-conserving surgery with whole-breast irradiation in stage I and II invasive breast cancer: American society of clinical oncology endorsement of the society of surgical oncology/American society for radiation oncology consensus guideline," *Journal of Clinical Oncology*, vol. 32, no. 14, pp. 1502–1506, May 2014, doi: 10.1200/JCO.2014.55.1572.

[6] M. S. Moran *et al.*, "Society of Surgical Oncology–American Society for Radiation Oncology Consensus Guideline on Margins for Breast-Conserving Surgery With Whole-Breast Irradiation in Stages I and II Invasive Breast Cancer," *International Journal of Radiation Oncology*Biology*Physics*, vol. 88, no. 3, pp. 553–564, Mar. 2014, doi: 10.1016/J.IJROBP.2013.11.012.

[7] F. Liewald, R. A. Hatz, H. Dienemann, and L. Sunder-Plassmann, "Importance of microscopic residual disease at the bronchial margin after resection for non-small-cell carcinoma of the lung," *The Journal of Thoracic and Cardiovascular Surgery*, vol. 104, no. 2, pp. 408–412, Aug. 1992, doi: 10.1016/S0022-5223(19)34796-8.

[8] O. Abe *et al.*, "Effects of radiotherapy and of differences in the extent of surgery for early breast cancer on local recurrence and 15-year survival: an overview of the randomised trials," *The Lancet*, vol. 366, no. 9503, pp. 2087–2106, Dec. 2005, doi: 10.1016/S0140-6736(05)67887-7.

[9] J. Ma, W. Li, J. Li, R. Shi, G. Yin, and R. Wang, "A small molecular pH-dependent fluorescent probe for cancer cell imaging in living cell," *Talanta*, vol. 182, pp. 464–469, May 2018, doi: 10.1016/J.TALANTA.2018.01.088.

[10] L. Liu *et al.*, "A near-infrared biothiol-specific fluorescent probe for cancer cell recognition," *Analyst*, vol. 144, no. 16, pp. 4750–4756, Aug. 2019, doi: 10.1039/C9AN00795D.

[11] S. Fu, X. Wan, C. Du, H. Wang, J. Zhou, and Z. Wang, "A novel fluorescent probe for the early detection of prostate cancer based on endogenous zinc sensing," *The Prostate*, vol. 79, no. 12, pp. 1406–1413, Sep. 2019, doi: 10.1002/PROS.23844.

[12] V. Sao, K. Pourrezaei, A. Akin, and H. Ayaz, "Breast tumor imaging using NIR led based handheld continuous-wave imager," *Proceedings of the IEEE Annual Northeast Bioengineering Conference, NEBEC*, pp. 55–56, 2003, doi: 10.1109/NEBC.2003.1215989.

[13] E. A. Kwizera *et al.*, “Molecular Detection and Analysis of Exosomes Using Surface-Enhanced Raman Scattering Gold Nanorods and a Miniaturized Device,” *Theranostics*, vol. 8, no. 10, p. 2722, 2018, doi: 10.7150/THNO.21358.

[14] O. T. Okusanya *et al.*, “Small Portable Interchangeable Imager of Fluorescence for Fluorescence Guided Surgery and Research:,” *http://dx.doi.org/10.7785/tcrt.2012.500400*, vol. 14, no. 2, pp. 213–220, Jan. 2013, doi: 10.7785/TCRT.2012.500400.

[15] K. Kerrou *et al.*, “The Usefulness of a Preoperative Compact Imager, a Hand-Held γ-Camera for Breast Cancer Sentinel Node Biopsy: Final Results of a Prospective Double-Blind, Clinical Study,” *Journal of Nuclear Medicine*, vol. 52, no. 9, pp. 1346–1353, Sep. 2011, doi: 10.2967/JNUMED.111.090464.

[16] M. Han, A. W. Partin, M. Zahurak, S. Piantadosi, J. I. Epstein, and P. C. Walsh, “Biochemical (prostate specific antigen) recurrence probability following radical prostatectomy for clinically localized prostate cancer,” *Journal of Urology*, vol. 169, no. 2, pp. 517–523, Feb. 2003, doi: 10.1016/S0022-5347(05)63946-8.

[17] L. E. McCahill *et al.*, “Variability in Reexcision Following Breast Conservation Surgery,” *JAMA*, vol. 307, no. 5, pp. 467–475, Feb. 2012, doi: 10.1001/JAMA.2012.43.

[18] N. Tagaya *et al.*, “Intraoperative identification of sentinel lymph nodes by near-infrared fluorescence imaging in patients with breast cancer,” *American Journal of Surgery*, vol. 195, no. 6, pp. 850–853, Jun. 2008, doi: 10.1016/J.AMJSURG.2007.02.032.

[19] E. M. Chan, E. S. Levy, and B. E. Cohen, “Rationally Designed Energy Transfer in Upconverting Nanoparticles,” *Advanced Materials*, vol. 27, no. 38, pp. 5753–5761, Oct. 2015, doi: 10.1002/ADMA.201500248.

[20] A. Teitelboim *et al.*, “Energy Transfer Networks within Upconverting Nanoparticles Are Complex Systems with Collective, Robust, and History-Dependent Dynamics,” *The Journal of Physical Chemistry C*, vol. 123, no. 4, pp. 2678–2689, Jan. 2019, doi: 10.1021/ACS.JPCC.9B00161.

[21] B. Tian *et al.*, “Low irradiance multiphoton imaging with alloyed lanthanide nanocrystals,” *Nature Communications 2018 9:1*, vol. 9, no. 1, pp. 1–8, Aug. 2018, doi: 10.1038/s41467-018-05577-8.

[22] E. P. Papageorgiou, B. E. Boser, and M. Anwar, “Chip-scale fluorescence imager for in vivo microscopic cancer detection,” *IEEE Symposium on VLSI Circuits, Digest of Technical Papers*, pp. C106–C107, Aug. 2017, doi: 10.23919/VLSIC.2017.8008565.

[23] E. P. Papageorgiou, B. E. Boser, and M. Anwar, “Chip-Scale Angle-Selective Imager for In Vivo Microscopic Cancer Detection,” *IEEE Trans Biomed Circuits Syst*, vol. 14, no. 1, pp. 91–103, Feb. 2020, doi: 10.1109/TBCAS.2019.2959278.

[24] E. P. Papageorgiou, B. E. Boser, and M. Anwar, “An angle-selective CMOS imager with on-chip micro-collimators for blur reduction in near-field cell imaging,” *Proceedings of the IEEE International Conference on Micro Electro Mechanical Systems (MEMS)*, vol. 2016-February, pp. 337–340, Feb. 2016, doi: 10.1109/MEMSYS.2016.7421629.

[25] M. Y. Berezin and S. Achilefu, “Fluorescence Lifetime Measurements and Biological Imaging,” *Chemical Reviews*, vol. 110, no. 5, pp. 2641–2684, May 2010, doi: 10.1021/CR900343Z.

[26] M. A. Mycek, K. T. Schomacker, and N. S. Nishioka, “Colonic polyp differentiation using time-resolved autofluorescence spectroscopy,” *Gastrointestinal Endoscopy*, vol. 48, no. 4, pp. 390–394, Oct. 1998, doi: 10.1016/S0016-5107(98)70009-4.

[27] J. Zhao, D. Zhong, and S. Zhou, “NIR-I-to-NIR-II fluorescent nanomaterials for biomedical imaging and cancer therapy,” *Journal of Materials Chemistry B*, vol. 6, no. 3, pp. 349–365, Jan. 2018, doi: 10.1039/C7TB02573D.

[28] A. D. Ostrowski *et al.*, "Controlled synthesis and single-particle imaging of bright, sub-10 nm lanthanide-doped upconverting nanocrystals," *ACS Nano*, vol. 6, no. 3, pp. 2686–2692, Mar. 2012, doi: 10.1021/NN3000737/SUPPL_FILE/NN3000737_SI_001.PDF.

[29] H. Najafiaghdam *et al.*, "A Molecular Imaging 'Skin A Time-resolving Intraoperative Imager for Microscopic Residual Cancer Detection Using Enhanced Upconverting Nanoparticles∗," *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, vol. 2018-July, pp. 3005–3008, Oct. 2018, doi: 10.1109/EMBC.2018.8512947.

[30] M. A. Green, J. Zhao, A. Wang, P. J. Reece, and M. Gal, "Efficient silicon light-emitting diodes," *Nature 2001 412:6849*, vol. 412, no. 6849, pp. 805–808, Aug. 2001, doi: 10.1038/35090539.

[31] R. L. Watters and G. W. Ludwig, "Measurement of Minority Carrier Lifetime in Silicon," *Journal of Applied Physics*, vol. 27, no. 5, p. 489, May 2004, doi: 10.1063/1.1722409.

[32] American National Standards Institute, *American National Standard For Safe Use Of Lasers*, ANSI Z136.1. New York, NY USA: Laser Institute of America, 2014.

[33] H. Najafiaghdam, E. Papageorgiou, N. A. Torquato, B. Tian, B. E. Cohen, and M. Anwar, "A 25 micron-thin microscope for imaging upconverting nanoparticles with NIR-I and NIR-II illumination," *Theranostics*, vol. 9, no. 26, p. 8239, 2019, doi: 10.7150/THNO.37672.

[34] Y. Maruyama, K. Sawada, H. Takao, and M. Ishida, "The fabrication of filter-less fluorescence detection sensor array using CMOS image sensor technique," *Sensors and Actuators A: Physical*, vol. 128, no. 1, pp. 66–70, Mar. 2006, doi: 10.1016/J.SNA.2006.01.017.

[35] R. K. Henderson *et al.*, "A 192 128 Time Correlated SPAD Image Sensor in 40-nm CMOS Technology," *IEEE Journal of Solid-State Circuits*, vol. 54, no. 7, pp. 1907–1916, Jul. 2019, doi: 10.1109/JSSC.2019.2905163.

[36] J. Koh and M. J. Kim, "Introduction of a New Staging System of Breast Cancer for Radiologists: An Emphasis on the Prognostic Stage," *Korean Journal of Radiology*, vol. 20, no. 1, p. 69, Jan. 2019, doi: 10.3348/KJR.2018.0231.

[37] Z. Li *et al.*, "A time-resolved CMOS image sensor with draining-only modulation pixels for fluorescence lifetime imaging," *IEEE Transactions on Electron Devices*, vol. 59, no. 10, pp. 2715–2722, 2012, doi: 10.1109/TED.2012.2209179.

[38] J. Guo and S. Sonkusale, "A 65 nm CMOS digital phase imager for time-resolved fluorescence imaging," *IEEE Journal of Solid-State Circuits*, vol. 47, no. 7, pp. 1731–1742, 2012, doi: 10.1109/JSSC.2012.2191335.

[39] T. SL *et al.*, "The FLARE intraoperative near-infrared fluorescence imaging system: a first-in-human clinical trial in breast cancer sentinel lymph node mapping," *Annals of surgical oncology*, vol. 16, no. 10, pp. 2943–2952, Oct. 2009, doi: 10.1245/S10434-009-0594-2.

[40] S. RM, C. NJ, and S. LP, "Near-infrared fluorescence imaging for real-time intraoperative anatomical guidance in minimally invasive surgery: a systematic review of the literature," *World journal of surgery*, vol. 39, no. 5, pp. 1069–1079, May 2015, doi: 10.1007/S00268-014-2911-6.

[41] de B. E *et al.*, "Optical innovations in surgery," *The British journal of surgery*, vol. 102, no. 2, Jan. 2015, doi: 10.1002/BJS.9713.

[42] R. Prevedel *et al.*, "Simultaneous whole-animal 3D imaging of neuronal activity using light-field microscopy," *Nature Methods*, vol. 11, no. 7, pp. 727–730, Jul. 2014, doi: 10.1038/nmeth.2964.

[43] N. C. Pégard, H.-Y. Liu, N. Antipa, M. Gerlock, H. Adesnik, and L. Waller, "Compressive light-field microscopy for 3D neural activity recording," *Optica, OPTICA*, vol. 3, no. 5, pp. 517–524, May 2016, doi: 10.1364/OPTICA.3.000517.

[44] M. Broxton *et al.*, "Wave optics theory and 3-D deconvolution for the light field microscope," *Opt. Express, OE*, vol. 21, no. 21, pp. 25418–25439, Oct. 2013, doi: 10.1364/OE.21.025418.

[45] N. Cohen *et al.*, "Enhancing the performance of the light field microscope using wavefront coding," *Opt. Express, OE*, vol. 22, no. 20, pp. 24817–24839, Oct. 2014, doi: 10.1364/OE.22.024817.

[46] N. Wagner *et al.*, "Instantaneous isotropic volumetric imaging of fast biological processes," *Nature Methods*, vol. 16, no. 6, pp. 497–500, Jun. 2019, doi: 10.1038/s41592-019-0393-z.

[47] R. Rabbani *et al.*, "Towards an Implantable Fluorescence Image Sensor for Real-Time Monitoring of Immune Response in Cancer Therapy," 2021.

[48] S. Ghosh and C. Preza, "Fluorescence microscopy point spread function model accounting for aberrations due to refractive index variability within a specimen," *Journal of Biomedical Optics*, vol. 20, no. 7, p. 075003, Jul. 2015, doi: 10.1117/1.jbo.20.7.075003.

[49] H. Wang *et al.*, "Deep learning enables cross-modality super-resolution in fluorescence microscopy," *Nature Methods 2018 16:1*, vol. 16, no. 1, pp. 103–110, Dec. 2018, doi: 10.1038/s41592-018-0239-0.

[50] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," Dec. 2015, [Online]. Available: https://arxiv.org/abs/1512.03385v1

[51] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: an overview and application in radiology," *Insights into Imaging*, vol. 9, no. 4. Springer Verlag, pp. 611–629, Aug. 01, 2018. doi: 10.1007/s13244-018-0639-9.

[52] T. J. Smith, A. Broome, D. Stanley, J. Westberg, G. Wysocki, and K. Sengupta, "A hybrid THz imaging system with a 100-Pixel CMOS Imager and a 3.25-3.50 THz quantum cascade laser frequency comb," *IEEE Solid-State Circuits Letters*, vol. 2, no. 9, pp. 151–154, Sep. 2019, doi: 10.1109/LSSC.2019.2933332.

[53] J. Choi *et al.*, "A 512-Pixel, 51-kHz-Frame-Rate, Dual-Shank, Lens-Less, Filter-Less Single-Photon Avalanche Diode CMOS Neural Imaging Probe," *IEEE Journal of Solid-State Circuits*, vol. 54, no. 11, pp. 2957–2968, Nov. 2019, doi: 10.1109/JSSC.2019.2941529.

[54] E. Rustami *et al.*, "Needle-Type Imager Sensor with Band-Pass Composite Emission Filter and Parallel Fiber-Coupled Laser Excitation," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 67, no. 4, pp. 1082–1091, Apr. 2020, doi: 10.1109/TCSI.2019.2959592.

[55] C. Lee, B. Johnson, T. Jung, and A. Molnar, "A 72 × 60 Angle-Sensitive SPAD Imaging Array for Lens-less FLIM," *Sensors (Basel, Switzerland)*, vol. 16, no. 9, Sep. 2016, doi: 10.3390/S16091422.

[56] A. Gharia, E. P. Papageorgiou, S. Giverts, C. Park, and M. Anwar, "Signal to Noise Ratio as a Cross-Platform Metric for Intraoperative Fluorescence Imaging," *Mol Imaging*, vol. 19, Apr. 2020, doi: 10.1177/1536012120913693.

[57] K. Perlin, "An image synthesizer," in *Proceedings of the 12th annual conference on Computer graphics and interactive techniques*, Jul. 1985, pp. 287–296. doi: 10.1145/325334.325247.

[58] A. Tatarinov, “Perlin noise in Real-time Computer Graphics.” 2008. [Online]. Available: /paper/Perlin-noise-in-Real-time-Computer-Graphics-Tatarinov/b49da45b19f6ad6c28b3748223b715810711d15f

[59] S. Zhao, H. Fu, M. Gong, and D. Tao, “Geometry-Aware Symmetric Domain Adaptation for Monocular Depth Estimation,” *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2019-June, pp. 9780–9790, Apr. 2019, Accessed: May 16, 2021. [Online]. Available: http://arxiv.org/abs/1904.01870

[60] F. Ma, G. V. Cavalheiro, and S. Karaman, “Self-supervised sparse-to-dense: Self-supervised depth completion from LiDAR and monocular camera,” in *Proceedings - IEEE International Conference on Robotics and Automation*, May 2019, vol. 2019-May, pp. 3288–3295. doi: 10.1109/ICRA.2019.8793637.

[61] J. M. Facil, B. Ummenhofer, H. Zhou, L. Montesano, T. Brox, and J. Civera, “CAM-Convs: Camera-Aware Multi-Scale Convolutions for Single-View Depth,” Apr. 2019, Accessed: May 16, 2021. [Online]. Available: http://arxiv.org/abs/1904.02028

[62] Y. Kuznietsov, J. Stückler, and B. Leibe, “Semi-Supervised Deep Learning for Monocular Depth Map Prediction,” *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, vol. 2017-January, pp. 2215–2223, Feb. 2017, Accessed: May 16, 2021. [Online]. Available: http://arxiv.org/abs/1702.02706

[63] F. Huang *et al.*, “Ultra-High Resolution 3D Imaging of Whole Cells,” *Cell*, vol. 166, no. 4, pp. 1028–1040, Aug. 2016, doi: 10.1016/j.cell.2016.06.016.

[64] “Fluorescent and Luminescent Probes for Biological Activity - 2nd Edition.” [Online]. Available: https://www.elsevier.com/books/fluorescent-and-luminescent-probes-for-biological-activity/mason/978-0-12-447836-7

[65] S. R. P. Pavani *et al.*, “Three-dimensional, single-molecule fluorescence imaging beyond the diffraction limit by using a double-helix point spread function,” *Proc Natl Acad Sci U S A*, vol. 106, no. 9, pp. 2995–2999, Mar. 2009, doi: 10.1073/pnas.0900245106.

[66] Z. Geng, “Depth Estimation from Single Image Using CNN-Residual Network,” 2017.

[67] C. Zhao, Q. Sun, C. Zhang, Y. Tang, and F. Qian, “Monocular Depth Estimation Based On Deep Learning: An Overview,” Mar. 2020, doi: 10.1007/s11431-020-1582-8.

[68] I. Laina, C. Rupprecht, V. Belagiannis, F. Tombari, and N. Navab, “Deeper Depth Prediction with Fully Convolutional Residual Networks,” in *2016 Fourth International Conference on 3D Vision (3DV)*, Oct. 2016, pp. 239–248. doi: 10.1109/3DV.2016.32.

[69] D. Eigen, C. Puhrsch, and R. Fergus, “Depth Map Prediction from a Single Image using a Multi-Scale Deep Network,” Jun. 2014, [Online]. Available: https://arxiv.org/abs/1406.2283v1

[70] I. Alhashim and P. Wonka, “High Quality Monocular Depth Estimation via Transfer Learning,” Dec. 2018, [Online]. Available: https://arxiv.org/abs/1812.11941v2

[71] M. Marcuzzo, P. Quelhas, M. Oliveira, A. M. Mendonça, and A. Campilho, “Cancer cell detection and invasion depth estimation in brightfield images,” 2009, [Online]. Available: https://repositorio-aberto.up.pt/handle/10216/76801

[72] D. Połap, M. Wozniak, M. Korytkowski, and R. Scherer, “Encoder-Decoder Based CNN Structure for Microscopic Image Identification,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12532 LNCS, pp. 301–312, Nov. 2020, doi: 10.1007/978-3-030-63830-6_26.

[73] X. Liu, S. Chen, L. Song, M. Woźniak, and S. Liu, “Self-attention negative feedback network for real-time image super-resolution,” *Journal of King Saud University - Computer and Information Sciences*, Jul. 2021, doi: 10.1016/J.JKSUCI.2021.07.014.

[74] L. li Guo and M. Woźniak, “An Image Super-Resolution Reconstruction Method with Single Frame Character Based on Wavelet Neural Network in Internet of Things,” *Mobile Networks and Applications*, vol. 26, no. 1, pp. 390–403, Feb. 2021, doi: 10.1007/S11036-020-01681-6/FIGURES/13.

[75] A. D. Newton, G. T. Kennedy, J. D. Predina, P. S. Low, and S. Singhal, “Intraoperative molecular imaging to identify lung adenocarcinomas,” *Journal of Thoracic Disease*, vol. 8, no. 9, Oct. 2016, doi: 10.21037/jtd.2016.09.50.

[76] W. S. Tummers *et al.*, “Intraoperative Pancreatic Cancer Detection using Tumor-Specific Multimodality Molecular Imaging,” *Ann Surg Oncol*, vol. 25, no. 7, pp. 1880–1888, Jul. 2018, doi: 10.1245/s10434-018-6453-2.

[77] Q. Zhou *et al.*, “Molecular imaging of a fluorescent antibody against epidermal growth factor receptor detects high-grade glioma,” *Scientific Reports 2021 11:1*, vol. 11, no. 1, pp. 1–12, Mar. 2021, doi: 10.1038/s41598-021-84831-4.