Hybridization process for back-illuminated silicon Geiger-mode avalanche photodiode arrays

Daniel R. Schuette, Richard C. Westhoff, Andrew H. Loomis, Douglas J. Young, Joseph S. Ciampi, Brian F. Aull, and Robert K. Reich

MIT Lincoln Laboratory, 244 Wood Street, Lexington MA 02420-9108

ABSTRACT

We present a unique hybridization process that permits high-performance back-illuminated silicon Geiger-mode avalanche photodiodes (GM-APDs) to be bonded to custom CMOS readout integrated circuits (ROICs) - a hybridization approach that enables independent optimization of the GM-APD arrays and the ROICs. The process includes oxide bonding of silicon GM-APD arrays to a transparent support substrate followed by indium bump bonding of this layer to a signal-processing ROIC. This hybrid detector approach can be used to fabricate imagers with high-fill-factor pixels and enhanced quantum efficiency in the near infrared as well as large-pixel-count, small-pixel-pitch arrays with pixel-level signal processing. In addition, the oxide bonding is compatible with high-temperature processing steps that can be used to lower dark current and improve optical response in the ultraviolet.

Keywords: Geiger-mode avalanche photodiode, back-illumination, oxide-bonding, photon-counting, imager

1. INTRODUCTION

Bandwidth requirements associated with processing a signal at the single-photon level are a challenge for single-photon detectors. In fast time-resolved applications traditional single-photon detectors, such as photomultiplier tubes and micro-channel plates have a relatively modest number of channels to process signals at the single-photon level. This introduces a limit on the instantaneous dynamic range [photons/sec] offered by these detectors. Increasingly, however, there is a need for fast time-correlated optical sensing, driven by applications in areas including communications, active imaging, and biology. This need has pushed the development of multi-channel arrays of single-photon sensitive detectors. While multi-channel versions of traditional technologies have been demonstrated, the most aggressive examples of scaling to higher array sizes leverages integrated circuit techniques to enable signal processing electronics required at the pixel level.

Recently, there have been a number of successful efforts to integrate single-photon-sensitive Geiger-mode avalanche photodiodes (GM-APDs) into commercial CMOS processes. These efforts have yielded devices with thousands to tens-of-thousands of channels, enabling a high degree of parallel signal processing that overcomes the instantaneous dynamic range limits of more conventional detectors. These gains, however, come with a substantial sacrifice in performance both because the detectors arrays share area with the pixel processing circuitry leading to poor fill factors and the detector architecture and fabrication processes are not optimized for optical performance leading to modest quantum efficiency (QE, < 50%) over a limited waveband (400 nm – 600 nm).

An alternative approach to integrating the GM-APD with the signal processing circuitry at the pixel level is fabrication of a hybrid detector where the signal processing electronics and the detector arrays are fabricated onto separate layers. An integration technique such as bump-bonding is used to connect detector elements with signal-processing pixels. Examples of this approach applied to GM-APD arrays are InP-based detectors produced by the Optoelectronic Materials group at MIT Lincoln Laboratory. These devices utilize flip-chip bump bonding to produce a hybrid detector where each detector element is directly connected to CMOS signal processing electronics, providing a degree of parallel processing that scales directly with the array size and enabling array sizes of thousands to millions of detector elements. These compound semiconductor single-photon detector arrays benefit from a transparent InP substrate that provides mechanical support during the flip-chip bonding process, making detector integration ultimately possible. While thick silicon photodiodes have been bump bonded, the relatively thin silicon GM-APDs needed for high-speed applications have not been amenable to this integration approach.
In the remainder of this paper, we will outline a silicon GM-APD hybrid detector fabrication method that combines a silicon GM-APD back-illumination process, wherein the thick silicon substrate of a GM-APD device wafer is replaced by a quartz substrate and GM-APD is integrated with the CMOS readout integrated circuits (ROIC) by flip-chip bump bonding. The final hybrid detector structure is illustrated in cross section in Figure 1 and will be described in detail in the following sections. The key advantage of this approach is the independent optimization of the GM-APD array and the ROIC, enabling detector performance that is optimized (in terms of fill factor, dark count rate, quantum efficiency, and detection timing) for a given application while making it possible to leverage commercial integrated circuit foundries to achieve the necessary miniaturization for very-high-element-count arrays. To-date, this work has focused on the active imaging applications of ladar and adaptive optics wavefront sensing.

2. SILICON GM-APD HYBRID DETECTOR ARCHITECTURE

Before presenting an overview of the GM-APD hybridization process a hybrid-detector architecture is described using, as an example, a GM-APD detector architecture that has been developed for ladar. This example architecture is similar to the design, described by Aull\textsuperscript{2}, in which the GM-APD utilizes a separate absorber-multiplier structure to collect photo-electrons from deep in the silicon bulk. The structure is a reach-through APD built in an epitaxial layer of silicon grown on a heavily doped silicon substrate. The GM-APD is formed by a p-type buried anode implant into the silicon with the dopant located at a small (sub-micron) separation from a surface cathode implant. A depletion region forms when a voltage is applied to the n\textsuperscript{+} cathode relative to the silicon bulk. As the voltage is increased this depletion region extends first to the anode implant then punches through extending the depletion region into the absorber. Because of the small separation between the anode and the cathode implants, the electric field in the multiplier is significantly higher than the absorber. By proper design of the buried implant dose and depth the multiplier electric field strength is optimized for high probability of avalanche breakdown and for depletion regions that extend through most of the silicon absorber. This results in high photon-detection-efficiencies (PDE = quantum efficiency \times avalanche breakdown probability, > 50\%) while collecting photoelectrons from a significant volume at modest breakdown overbias levels (< 6 V). If the light enters from the cathode-side (also referred to as the front side) of the GM-APD photons are blocked by the metal interconnects used to electronically access the detector. As a result, the scalability of this architecture to large-element-count arrays is limited and also the optical path into the photosensitive regions has a number of interfaces that cause reflections and absorbing layers that lead to losses, particularly at short wavelengths.

As depicted in Figure 2, the detector can be back-illuminated by removing the heavily doped silicon substrate. In this structure, a photon enters the detector from the back side (anode side) of the devices, opposite the metal interconnects, and is then converted into a photo-electron in the absorber region. As with the front-side GM-APDs, reach through fields in the absorber draw the photo-electron into the multiplier region where it initiates an avalanche breakdown. Because of the absorber region can be tens of microns thick this structure can have good quantum efficiency into the near infrared. In addition, using backside passivation techniques developed for scientific imagers (\textit{e.g.} molecular-beam epitaxy-MBE), it is possible to extend the performance of the detector deep into the ultraviolet. In both front- and back-side devices, the electric field within the multiplier region primarily determines the avalanche breakdown probability. Since the electric field extends through most of the absorber region, high photon-detection efficiencies can be obtained over a broad band of wavelengths at modest overbias levels.
Figure 2: Example cross section of a back-illuminated GM-APD device. For a measure of scale, the GM-APD device layer is typically between 5 and 20 µm thick (depending on device architecture) while the diameter of the anode diameter ranges from 20 µm to 50 µm. Light enters through the top of the quartz substrate and is converted to photoelectrons in the absorber. The anti-reflection coating and backside passivation are used to optimize the optical interface between the quartz and silicon.

Electrical connection to the GM-APD is made with a bump bond directly between the GM-APD cathode and a signal-processing pixel on the CMOS ROIC. The electrical connection to the substrate is made through a ring of bump bonds around the edge of the GM-APD array which biases the silicon bulk and the GM-APD anode. Bump bonding enables a (high) density of interconnects that distinguishes this detector from other GM-APD arrays that use optimized processes to improve detector performance (e.g. commercial devices offered by Hamamatsu, PerkinElmer, and others). These commercial detectors are small, typically linear, arrays with less than one-hundred elements. In contrast, present bump bonding technology supports hybridization of very-large-element-count detector arrays, with routine fabrication of bump-bonded hybrid imagers with tens-of-thousands of pixels and commercial examples of multi-mega-pixel devices (e.g. the 4k × 4k hybrid imagers produced by Teledyne Scientific Imaging). With this hybridization process, it is possible to fabricate single-photon-sensitive arrays with element-count, pixel-pitch and pixel-size comparable to scientific imagers (the example you give in Figure 1 make it seem like the pixels are big).

While the example GM-APD architecture discussed above is for photon-timing applications, such as 3-D imaging ladar, the back-illumination process presented in the following section is versatile and can be applied to other GM-APD structures as well (e.g. photon-counting detectors such as the 100% fill-factor quad-cell array developed for adaptive optics wavefront sensors reported in3). Using the hybridization approach, a variety of GM-APD detector architectures can be fabricated on a wafer and integrated to existing CMOS readouts that use different foundry process technologies. This enables efficient and flexible use of the GM-APDs and CMOS ROICs in support of many applications (e.g. ladar, adaptive optics, communications, etc.). Basic architectures, already mentioned, are photon-timing and photon-counting. These basic architectures can have variations that are optimized for particular applications and include features like coincidence measurements and synchronous or asynchronous arming of the GM-APD detectors. By placing the detector on a separate tier, modified circuit architectures may be realized without affecting the performance of the detector array. In addition, as integrated circuit technology advances to higher-density, higher-performance fabrication nodes, the two tier approach allows the CMOS ROIC to advance without necessitating redesign of the detector array. Finally, the pixel pitch achievable with two-tier hybridization should have an advantage over the monolithic approach, because it avoids detector fill-factor trade-offs and can use the finer-feature-size processes offered by CMOS foundries.
3. GM-APD HYBRIDIZATION PROCESS OVERVIEW

As summarized in Figure 3, the process for fabricating hybrid detectors with back-illuminated silicon GM-APD detector arrays has six major steps that are described below. The process sequence begins after GM-APD arrays have been fabricated through the front-side processing, which includes device cathode and buried anode implants and metal layers. The silicon starting material may either be an epitaxial or a high-resistivity silicon wafer and the device architecture can be optimized for specific applications such as photon timing or photon counting, as previously discussed.

3.1. Silicon-handle wafer oxide bonding

GM-APD devices are presently processed on 150-mm wafers. The final thinned GM-APD device layer is typically less than 20-μm thick and is too fragile to handle freestanding so the GM-APD wafer is bonded to a temporary support that will be removed at a later stage in the process. To provide this support a silicon-handle wafer is attached to the front side (cathode side) of the GM-APD device wafer. Net-stress in the device and handle wafers must be limited so that the bow in each wafer is a few microns or less. Oxide layers are then deposited and planarized on both the GM-APD device wafer and the silicon-handle wafer to produce surfaces with local average smoothness of angstroms. When brought in contact, the activated oxide surfaces initiate a bond that is then further strengthened through a thermal anneal at 175 C for three hours. During this anneal it is important that gasses, primarily water vapor, that evolve at the bond interface be able to diffuse out of the structure. Due to the thickness of the handle-wafer and unthinned device wafer these gasses cannot escape if the bonding interfaces are strictly planar; therefore, lateral venting trenches are etched into the bonding oxide to provide the requisite diffusion path. With the oxide-bonded silicon-handle wafer, even a GM-APD device layer that is only a few microns thick can be processed with standard microelectronic fabrication tools.
3.2. Device wafer thinning

Because the final device will be illuminated through the back side of the wafer, it is important that photoelectrons are collected from the full thickness of GM-APD device absorber layer for high quantum efficiency and good timing resolution. The final detector layer thickness is a trade-off between detector response time and photon-detection efficiency at long wavelengths. Depending on the needs of the particular application, GM-APD device wafers supported by silicon-handle wafers are thinned to a thickness of few microns to tens of microns. Operationally, this is done by first grinding off approximately 600 microns of silicon and then using a combination of wet etching and chemical mechanical polishing to generate a defect-free surface with the desired thickness.

3.3. Backside processing

Backside processing of the GM-APD device layer is important for ensuring the high optical performance of back-illuminated devices as compared to front-illuminated devices. The passivation and anti-reflection coating steps optimize the optical interface through minimization of reflection and absorption losses. Reflection losses are minimized at the quartz-to-silicon and quartz-to-air interfaces by application of an anti-reflection-coating. Absorption losses are minimized by adapting back-side processing techniques developed for back-illuminated CCD imagers. For devices needing high quantum efficiency in the ultraviolet, the thinned GM-APD device wafer is passivated by growing a layer of highly boron doped silicon that is only a few nanometers thick, a delta-doping\footnote{\textsuperscript{10}}, on the back side of the device using MBE. This layer is used to passivate the back surface of the device, which reduces dark counts as well as produces an electric field within the device that drives photo-electrons toward the multiplier region of the GM-APD. Ion-implantation with laser-annealing is also used, in place of MBE, where short wavelength response is not as important. After backside passivation, metal layers that can act as light shields or return current paths and optical coatings that prevent reflection losses at the quartz-to-silicon interface may be added.
3.4. Quartz wafer oxide bonding

Quartz (fused silica) is chosen as a final handle wafer because of its compatibility with the oxide-bonding process and high transparency from ultraviolet to near-infrared wavelengths. In the fourth major step of the process, the quartz-handle wafer is attached to the backside of the thinned GM-APD device wafer by depositing an oxide layer on both the GM-APD device wafer and the quartz-handle wafer, then planarizing these layers until local roughness is on the order of angstroms. As with the silicon-handle wafer, when brought into contact the activated surfaces of these oxide layers initiate a bond. Strengthening this bond is more complicated than in conventional silicon-to-silicon wafer bonding due to the thermal mismatch of silicon and quartz. Specifically, it has been found that annealing at 125°C for approximately three hours is sufficient to remove residual water from the bond interface, critical in preventing the development of post-anneal voids, but low enough to prevent slipping of the bond surface during the anneal, which would generate unacceptable post-anneal wafer stress. As with the bonding of the silicon handle wafer (section 3.1) venting trenches are etched into the bonding oxide to prevent the formation of voids due to outgassing.

3.5. Silicon-handle wafer removal and final front-side processing

Following oxide-bonding of the quartz handle wafer, the thin GM-APD device layer is held between the quartz and silicon wafers. The silicon handle wafer is removed with grinding and selective chemical etching, a process similar to that used in wafer thinning. The bonding oxide provides a convenient etch stop so the process is easier. After the handle wafer is removed, the only silicon remaining is the thin active GM-APD device layer, now supported by a quartz substrate. The final front-side processing is completed by etching the oxide from the APD bump-bond pads in preparation for flip-chip hybridization. A photograph of a completed back-illuminated GM-APD wafer is shown in Figure 4.

3.6. Bump-bonding

To fabricate the final hybrid detector an electroless plating process, called zincation electroless nickel gold (ZENIG)\(^\text{11}\), is used to grow an under-bump metal on the CMOS readout integrate circuit. The ZENIG process avoids plasma deposited metal processes that can damage the CMOS. Also the ZENIG process enables under bump metallization of small chips that otherwise would require lithographic processing, which is difficult on the single chip scale. The GM-APDs have Ti/Ni/Au under bump metals deposited on the APD pad opening followed by the evaporation of 7-μm high indium bumps. A photoresist lift-off process is used to pattern the bumps. Detector arrays, diced out of the quartz-oxide-bonded...
GM-APD wafer, are then flip-chip bonded to the CMOS circuitry. Finally, an epoxy is used to underfill the GM-APD array and thereby ensure mechanical reliability. Figure 5 shows photographs of a completed hybrid detector.

CONCLUSIONS

Large-element-count GM-APD arrays have been fabricated using a back-illumination process that enables independent optimization of the detector and CMOS ROIC. In the case of the detector, optimization occurs on two fronts: first, the optical interface may be optimized to minimize light lost in the path to the detectors active region, through high-performance back side processing techniques (developed for solid-state imaging system such as CCDs and active pixel sensors) and application of anti-reflective coating; and second, the size and depth of the photo-collecting region may be structured to optimize for trade-offs between timing accuracy, dark count rate, and long-wavelength quantum efficiency. In the case of the signal processing circuit, the 3-D hybrid integration removes trade-offs between detector fill-factor and pixel circuit area found in monolithic commercial CMOS foundry-based approaches.

To-date, this hybridization process has been targeted at the active imaging applications of 3-D ladar imaging and adaptive optics wavefront sensing. Given, however, the versatility of the approach and the potential for extended detector response at wavelengths from into the deep ultraviolet to the near-infrared we anticipate that this technology offers opportunities for applications in other fields as well.

ACKNOWLEDGEMENTS

This work was sponsored by the Air Force under the Department of the Air Force contract number FA8721-05-C-0002. Opinions, interpretations, conclusions, and recommendations are those of the author and are not necessarily endorsed by the United States Government.

REFERENCES