Reduction of Band-to-band Tunneling in Deep-submicron CMOS Single Photon Avalanche Photodiodes

Robert K. Henderson[1], Justin Richardson[1,2], Lindsay Grant[2]
Robert.Henderson@ed.ac.uk, Tel: +44 131 650 5568
Justin.Richardson@st.com, Lindsay.Grant@st.com, Tel: +44 131 336 6000

SUMMARY
We demonstrate that a process layer intended for pinned-photodiode formation can be employed to reduce the dark count rate (DCR) of single photon avalanche diodes (SPADs) in nanometer-era CMOS technologies. A low-doped p-type passivation implant reduces the electric field in the p-diffusion/nwell breakdown region and acts both as a guard ring and as an STI edge interface. An 8µm diameter device with a typical DCR of 40Hz at 25C is reported in a 130nm CMOS imaging process.

I. INTRODUCTION
Single-photon avalanche diodes operating in Geiger mode are highly promising solid-state replacements for photomultiplier tubes (PMTs) in scientific or biomedical imaging [1,2]. Integration of these detectors in CMOS technology is a relatively recent innovation with the first detectors reported in 0.8µm CMOS, progressing rapidly in recent years to deep-submicron processes [3,4].

A number of CMOS SPADs have been proposed recently at 180nm or 130nm process nodes [5,6,7,8,9]. However, the authors consistently report excessive noise levels induced by band-to-band tunneling in the high field region of the device. Suitable field profiles for the SPAD avalanche region cannot easily be obtained due to the increased doping levels of source/drain p-diffusion and n-well in scaled CMOS process technologies. Thus, trends for PMOS transistor formation are now contrary to the requirements for SPAD detectors. Yet nanometer-era digital processing is an enabling technology for future single-photon imagers, promising picosecond time-to-digital conversion and large, low-power, time-resolved pixel arrays in a small form-factor.

In this paper we demonstrate that a single process implant found in CMOS image sensor technology can be employed to reduce the dark count noise in p-diffusion/n-well SPAD structures by several orders of magnitude. The detector can be fabricated without additional mask steps or doping level modifications.

II. DEEP SUBMICRON CMOS SPAD DEVICE STRUCTURES
A number of SPAD structures have been proposed in deep-submicron CMOS in recent years (Figures 2a-e) [3,4,5,6,7,8,9]. A SPAD with low DCR and high photon detection efficiency (PDE) requires the engineering of (1) a suitable guard ring to prevent premature breakdown at the device periphery and (2) a uniform, high field, breakdown region. While a number of guard ring designs have been proposed, the breakdown region is consistently formed by the p-n junction between p+ diffusion source-drain implant and n-well.

At 0.8µm and 0.35µm process nodes, SPADs with implicit guard rings, formed by either an enrichment n-well implant or by n-well spacing, were successful at providing low DCR [1,3,4]. Beyond the 250nm node, the introduction of STI and triple-well process steps has required new SPAD structures and we focus mainly on these modern process generations in this paper.

Fig. 1 180nm CMOS SPAD structures (a) [7] (b) [8]

The most common approach is to create an explicit guard ring of low doped p-well material around the central p+/n-well breakdown region. Two SPADs with such a structure have been reported at the 180nm process node [7,8] with the former at around 100Hz DCR and the latter at around 100kHz. Both authors cite band-to-band tunneling effects as being present. Yet another SPAD at 180nm [9], proposes the use of STI as the guard region but with even higher DCR of 1MHz thought to be due to traps at the STI-interface.

At the 130nm process node, various devices employing a p-well guard ring have been reported, invariably with high DCR (40-100kHz). Passivation of STI-interface traps has failed to reduce DCR significantly leading to the assumption that tunneling is the dominant mechanism. This is evidenced by DCR temperature dependence, and confirmed by TCAD modelling showing that the high field in the narrow p+/nwell depletion region is giving rise to excessive band-to-band tunneling [10]. Unfortunately, this is a trend that is set to continue, as doping levels are being increased in nanometer-era CMOS to maintain transistor performance.
III. NEW SPAD DEVICE STRUCTURE

It is clear from the previous discussion that the p+/nwell breakdown region must be modified to reduce the field strength. There are various possible approaches (1) change the doping profile by process modification (2) choose entirely different, lower-doped p or n layers (3) modify the doping profile of the existing p+/nwell junction by an additional implant. Approaches (1) and (2) have been shown to be successful in forthcoming publications [10][11]. In this work, we demonstrate that strategy (3) can also be applied.

The SPAD structure proposed in this paper is an adaptation of an enrichment SPAD structure [1][4] shown in Fig. 3 to a triple-well technology. The guard ring is formed implicitly, by the absence of p-well, creating a ring of p-epi doped with the graded (retrograde) doping profile of the buried n-well. The p-well formation can be prevented by a drawn ring of implant stop layer. The high field at the interface between n-well and p+ diffusion will favour breakdown in this planar region without, however, solving the band-to-band tunneling problem.

The cross-section of the new device structure is shown in Fig. 4 and a micrograph is shown in Fig. 5. The cathode terminal is formed by a deep retrograde n-well with the n-well enrichment implant defining the active breakdown region. The deep n-well is contacted at the periphery of the device by a ring of n-well with n-diffusion contacts.

IV. RESULTS

The anode terminal is formed by the conjunction of a conventional p-diffusion ring surrounded by a deeper, low-doped p-type implant. This p-implant is available in the CMOS imaging process where it fulfils the role of a glove-like passivation implant around the STI to reduce dark current in pinned-photodiodes, as employed originally in Fig. 2b [6]. The combination of the two p-type implants forms a less abrupt, graded junction which greatly reduces tunneling dark count compared to the conventional p+/n-well junction. The p-implant also acts as a guard ring structure. Whereas conventional guard rings are formed by a ring of lower doped p-well, we prevent p-well formation by a drawn implant stop layer. Instead, the lower field region is formed in the ring where the p-implant extends beyond n-well. This region is also where the p-epi wafer material is implanted with retrograde deep n-well. The p-implant makes a natural interface to the STI edge without the intermediary of a polysilicon thin oxide ring as in [5]. STI-bounded SPADs are highly attractive for dense integration with electronics [9].

Fig. 6 shows the IV curve of the device. The breakdown voltage is around 12.4V indicating a lower field in the avalanche region when compared to 9.6V breakdown of p-diffusion/n-well SPADs [5,6,7,8]. The current at breakdown is 124.6pA.
peak of around 20% at 450nm increasing with excess bias voltage. The PDE is improved by the reduced stack height of the imaging process. Finally, a plot of device jitter is given in Fig. 9 showing the absence of a significant tail and an FWHM of around 200ps. A 470nm, 80ps jitter pulsed laser diode has been employed in measurements.

V. CONCLUSIONS

Use of the extended palette of layers intended for pinned-photodiode formation in CMOS imaging processes has been effective at reducing dark count of a p+/nwell SPAD structure. Furthermore, we believe that a low noise device, dispensing with the p- ring and relying solely on the STI boundary will be possible since the tunnelling-induced DCR has been addressed by the p+/p-/n-well graded junction.

REFERENCES

Fig. 8 Photon detection efficiency versus wavelength at various excess bias voltages

Fig. 9 Jitter at various excess bias voltage with illumination from a pulsed 470nm laser diode