# PROCEEDINGS OF SPIE

# A theory for image persistence in HgCdTe photodiodes

Roger M. Smith, Maximilian Zavodny, Gustavo Rahmer, Marco Bonati

**SPIE.**

# A theory for image persistence in HgCdTe photodiodes

Roger M. Smith[*a], Maximilian Zavodny[a], Gustavo Rahmer[a], Marco Bonati[b]
[a]Caltech Optical Observatories, California Institute of Technology,
MC105-24, 1200 E. California Blvd, CA 91125, USA
[b]Cerro Tololo Inter-American Observatory, Casilla 603, La Serena, Chile.

## ABSTRACT

Image persistence, an elevated dark current caused by prior illumination, has been a chronic problem in near infrared focal plane arrays. It has the potential to interfere with astronomical surveys requiring precision photometry, astrometry and/or shape measurement. We posit a mechanism whereby charge is captured by traps exposed to carriers as the depletion width shrinks when photo-generated charge accumulates on the electrically isolated photodiode. After the depletion width is reestablished by resetting the diode, trapped charge is slowly released, appearing to be a signal during the subsequent exposure(s). We present data illustrating the complex behavior of persistence for several 1.7 μm cutoff HgCdTe arrays on Teledyne H2RG multiplexers, and show that our simple model predicts this behavior at least qualitatively. We identify one inconsistency that has yet to be resolved and discuss mitigation options.

**Keywords:** persistence, remnant image, HgCdTe, 1.7 μm cutoff, depletion region, charge traps

## 1. INTRODUCTION

The behavior of image persistence, a subtle increase in dark current in response to prior illumination, is complicated, and the authors were not aware of even a good qualitative model prior to our investigation. This has led to several widespread misconceptions or half-truths such as "*persistence is a response to saturation*" or "*remnant image can be erased by additional resets*".

In this paper we present a hypothesis for the persistence mechanism. Information on dopant and trap density profiles required for a quantitative analysis is not available, nevertheless the complexity of persistence behavior provides considerable evidence to support of our qualitative model. In a companion paper on persistence calibration[1], we discuss this complex behavior in more detail and describe the steps required to correct for the effect.

The model to be presented is based on the behavior of HgCdTe photodiode arrays, which were selected for low persistence. Additional phenomena may come in to play in devices with inferior performance. In particular, some devices have been reported to exhibit a signal threshold at which persistence increases sharply. At least in the case of the well-behaved devices that we tested, image persistence increases monotonically with stimulus amplitude and eventually saturates (Fig. 4) when the stimulus drives the diode into forward bias. Persistence is *not* a consequence of saturation. The response exhibits multiple time constants, and those time constants depend on the duration of the stimulus[1]. Persistence ratio expressed as a fraction of stimulus amplitude varies from pixel to pixel, and can vary significantly among detectors.

Given this complicated behavior and particularly its dependence on the exposure cadence and readout timing[1], performance projections from test data taken by others will be misleading unless the test method is agreed upon and, preferably, mimics the planned exposure cadence and sampling scheme. If the persistence quoted by the manufacturer is based on a short bright exposure, it will significantly underestimate the effect seen for the long exposures more typical of spectroscopy or imaging in space.

Recent HgCdTe focal plane arrays produced by Teledyne Imaging Systems for the SuperNova Acceleration Probe (SNAP) program[2]. with 1.7μm cut-off wavelength (HAWAII-2RG serial numbers 102 and 103) exhibit a remnant image in the following dark frame which is 0.23% and 0.46% (respectively) of the value in the previous frame, when observing at the SNAP survey cadence: 300s exposures, separated by 30s with shutter closed, during which time 11 non-

---

[*] rsmith@astro.caltech.edu; phone 1 626 395-8780; fax 1 626 568-1517

destructive reads are performed at beginning and end of exposure separated by one reset scan, at 1.3s per frame scan. Persistence decays to the level of the read noise (10e for Fowler-11 sampling⁻) by the 4th subsequent dark frame, if the stimulus was unsaturated. However, persistence below the noise level can be relevant for observing programs which are limited by systematic errors rather than noise.

All persistence data to be presented here represent the *excess* after subtraction of the dark signal measured for matched observing conditions (exposure duration, cadence, and sampling algorithm) when the detector has been operated at stable voltage bias and temperature for long periods with no prior illumination for several hours.

## 2. PERSISTENCE MODEL

We posit that image persistence is a charge capture and emission process occurring in those regions of the photodiode PN junction, which are intermittently exposed to carriers as the width of the depletion region is reduced by the accumulation of photo-generated charge (Fig. 1). Trapped charge remains behind when the depletion edge recedes as the diode is reset to its original reverse bias. Resetting the diode continuously or repeatedly has no further effect on the depletion width and thus does not affect the rate of charge emission from traps. It will however serve to delay the start of the next exposure so that some charge will be released before the first read that establishes the signal baseline.

A significant number of traps have emission time constant exceeding the time from beginning of Reset to the sampling of the baseline for the next exposure: charge released by traps after the baseline sample appear as a dark signal in the next frame. Charge may continue to be emitted for several more frames.
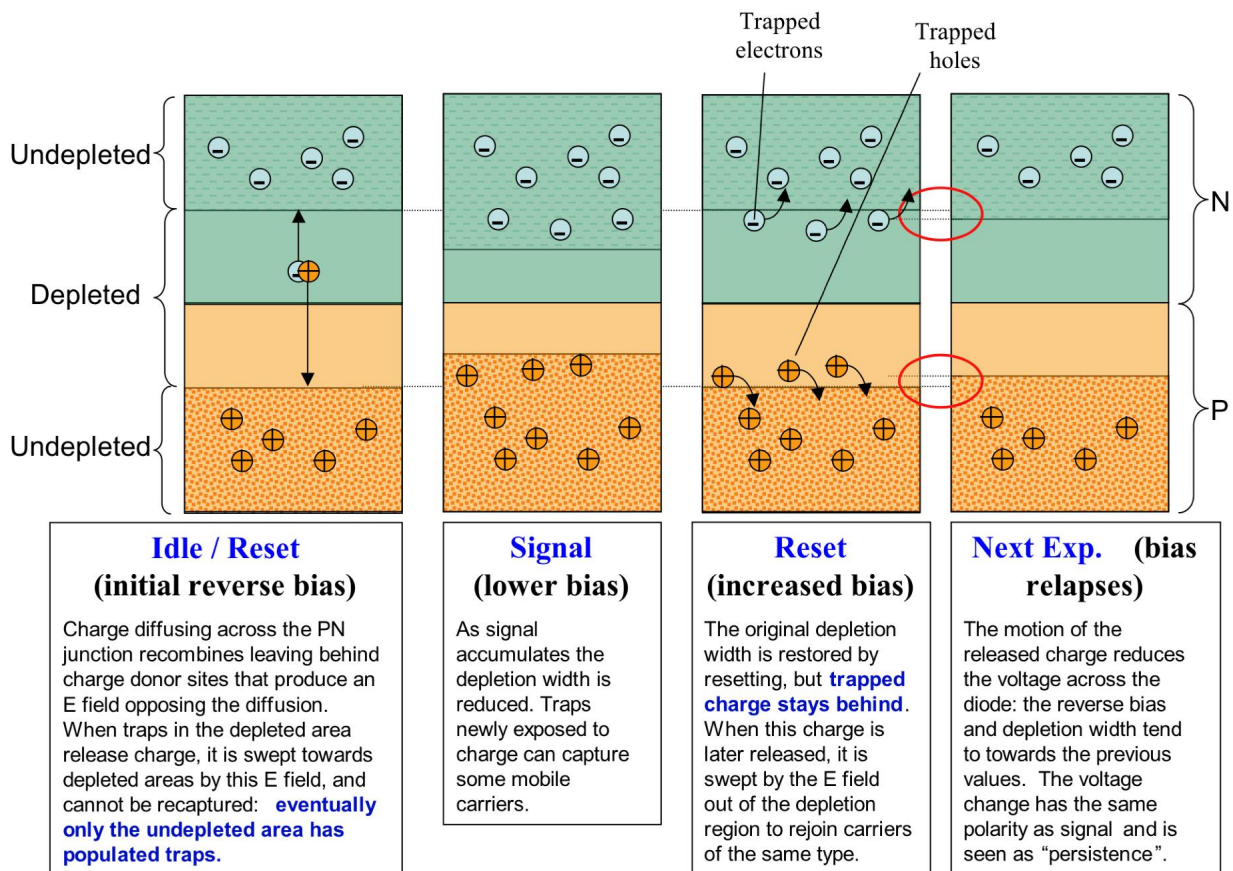


Fig. 1: Persistence model: as photo-generated charge accumulates, the depletion region shrinks exposing traps at the edge of the depletion region to carriers. Trapped charge begins to be released once the depletion edge recedes (when the diode is reset) and typically continues to be released for up to ten times the exposure duration.

Fig. 2 and Fig. 3 show the relationship between charge density profile in the depletion region, electric field (the integral of enclosed charge) and the voltage (the line integral through the electric field). The initial equilibrium condition is established as majority carriers thermally diffuse to the opposite side of the PN junction where they become minority carriers and rapidly recombine. This recombination leaves a static charge on the donor/receptor atoms which produces an electric field that opposed further charge diffusion. The voltage across the diode is always the line integral of the electric field along any path across the depletion region, so the recombination, which creates the depletion region also generates a "built in potential". Increasing the reverse bias (resetting the diode) removes charge from the depletion region and increases its width. As electron hole pairs are created by photons, they are separated by the electric field and flow to the outer region of the depletion region where they neutralize some of the charged donor/receptor atoms. The depletion width is reduced and the associated voltage drop is registered as a signal. The rate at which the depletion width collapses depends on flux and the dopant density profile.
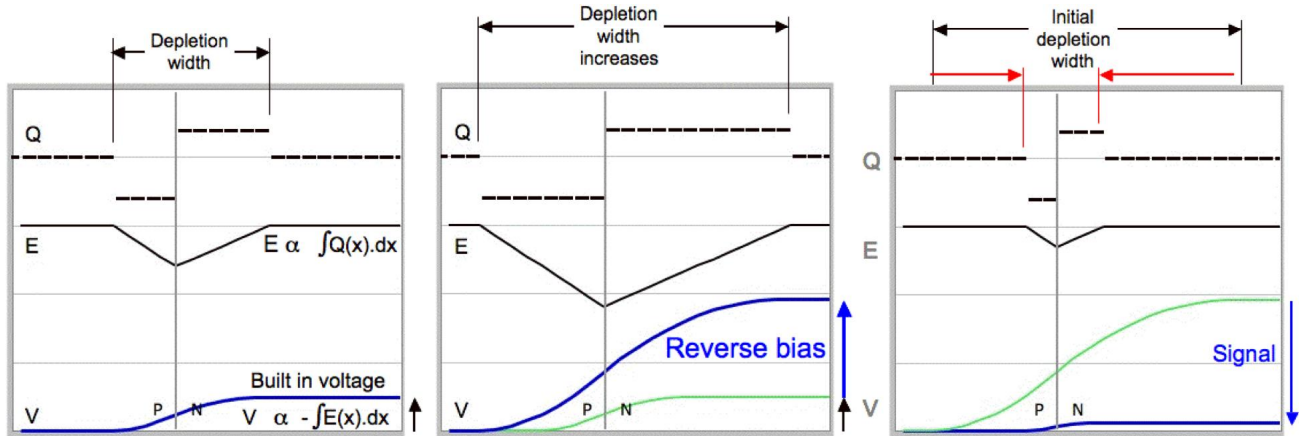


Fig. 2: Left panel: mobile carriers diffusing across the PN junction recombine to produce charged donor/receptor atoms whose electric field resists further thermally driven charge diffusion. The electric field is proportional to the integral of the enclosed charge and the associated "built in" voltage is proportional to the line integral through the electric field. When a reverse bias is applied (center panel) charge is drawn out of the depletion region, which widens. The same integrals define the relation ship between donor density profile, depletion width, E field and voltage. As electron hole pairs accumulate, they are separated by the E field, until they neutralize donor/receptor sites at the edge of the depletion region. The depletion width and the voltage across the diode are reduced. (right panel).
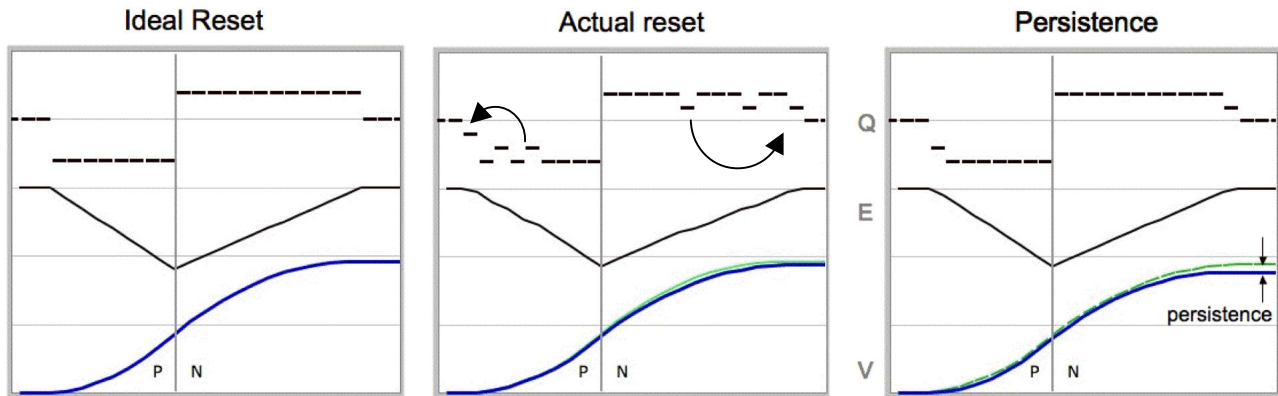


Fig. 3: The closure of the reset switch reestablishes the reverse bias voltage but instead of the original charge distribution (left here or center in Fig. 2), charge remains behind in traps, effectively lowering the charge density in the junction. To produce the same voltage, the depletion region is slightly wider. The subsequent emission of the trapped charge (right) leads to a reduction in reverse bias, which is mistaken for a signal.

When the diode is again reverse biased (reset) carriers which have been trapped remain behind. Their eventual release allows them to move under the influence of the electric field in the depletion region causing a small voltage drop, which

is interpreted as a signal in subsequent frames. It is not the presence of trapped charge which causes the spurious signal but the motion of these charges once released.

# 3. EVIDENCE FOR MODEL

This model is supported by the following observations:

- Persistence increases monotonically with stimulus then saturates when the diode reaches forward bias and the depletion width changes little (Fig. 4).

- Subdividing exposures reduces persistence considerably: less volume of the depletion region, and thus fewer traps, are exposed to charge (Fig. 5).

- Continuous reset during illumination suppresses persistence, since no (new) traps are exposed to charge when there is no change in depletion width.

- Persistence is higher if the stimulus is delivered as a flash at exposure start instead of a constant flux integrated to the same level (Fig. 4). Traps are exposed to charge for longer on average and thus capture more charge.

- Optical and electrically induced changes in detector bias produce similar persistence, provided that a step change in voltage is compared to a short flash of light, or a voltage ramp is compared to constant flux (Fig. 6, Fig. 7).

- Gain (volts per electron) is much lower for persistence than for photo-generated charge. Gain for persistence increases with stimulus amplitude (whereas gain for photo-generated charge is independent of charge absorption location). This surprising effect is predicted by the model (see section 3.4 below).

An alternative explanation sometimes put forward for remnant images is that the reset is incomplete or leaky. While the proportionality of the remnant image brightness to the stimulus intensity (Fig. 4) might be expected for incomplete reset, this should be manifested immediately and be dealt with by subtraction of the sample recorded at the beginning of the exposure, but continuous non-destructive readout shows that charge is emitted continuously well beyond the time at which the baseline is established.

A leaky reset switch is precluded as an explanation for persistence since this would produce a dark current dependence on signal in the current frame not on signal in the previous frame.

## 3.1 Persistence monotonicity

Persistence observed in the subsequent (dark) frame increases monotonically with the stimulus brightness, since the number of traps exposed to carriers increases as the edge of the depletion region moves towards the PN junction.

Over the majority of the range where signal is linear the persistence is also linear, so image persistence can be described as a simple percentage. This percentage varies from pixel to pixel and with time since stimulus. Above ~80,000 e- the persistence increases more quickly, possibly due to the dependence of voltage gain on trap distance from the depletion edge (described in section 3.4).

As the diode approaches forward bias, the repulsive electric field drops and the depletion width narrows allowing increasing minority carrier diffusion across the PN junction. This forward conduction (and associated recombination with photo-generated charge) increasingly cancels the photocurrent so the rate of change of the depletion width due to signal integration slows dramatically and fewer traps capture charge, so persistence saturates.

Comparing signal delivered in a brief flash at the start of the exposure to the same signal delivered as a constant flux provides a useful test of the persistence model in Fig. 1. Persistence is greater per unit stimulus for the flash since the entire volume sees charge for the whole exposure instead of being filled progressively. The saturation is sharper and lower since the leakage current discharges the signal and reestablishes sufficient depletion width for the leakage current to again become negligible. This places a lower and sharper limit on the volume in which traps see charge.

The curves in Fig. 4 show fluences on the X-axis, which greatly exceed the range of the AD converter. These values were determined by calibrating the flux from the IR LED against current at a shorter exposure time.
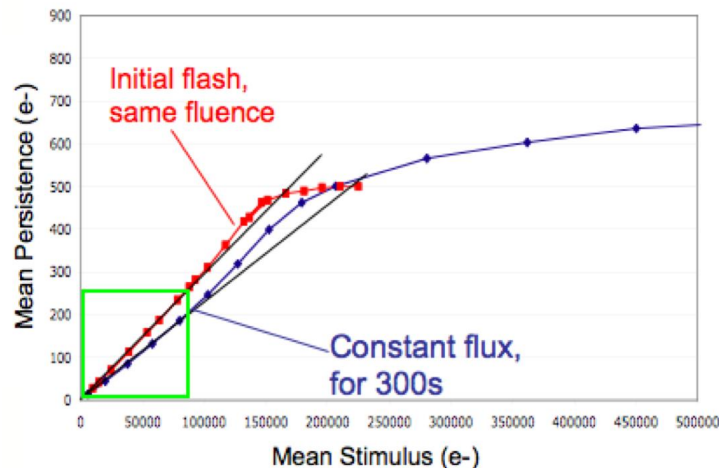
Fig. 4: Mean persistence in the first 300s dark after a 300s stimulus exposure with single reset frame in between. Persistence is linear with stimulus amplitude in the normal signal range (box) and can thus be described as a simple percentage. The lower curve is representative of normal exposures where flux is constant, while the upper curve is for a bright flash at the start of the exposure.

## 3.2 Exposure subdivision

Fig. 5 shows that when an exposure is subdivided less persistence occurs. This is to be expected since a smaller volume of the depletion region is exposed to charge, so that fewer traps are populated. This mitigation of persistence scales slightly faster than 1/N when subdividing into N sub-exposures, possibly because there are a decreasing number of unpopulated traps available in the latter sub-exposures.
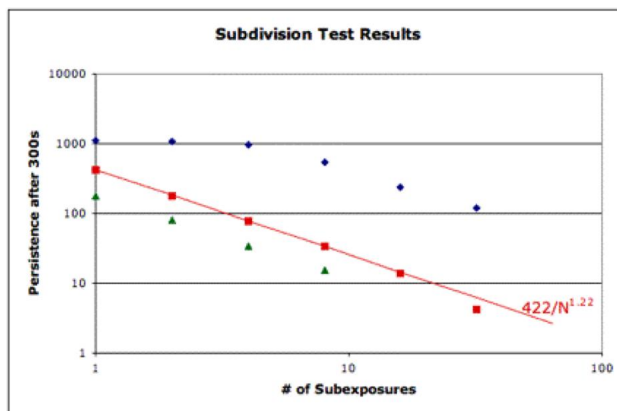


Fig. 5: When an exposure is subdivided charge accumulates repeatedly in a thinner layer of the depletion region so that fewer traps are exposed intermittently to charge. Plots for three different fluxes are shown. The first three points on the upper curve show no reduction in persistence since the diodes are driven into forward bias (saturation) even when exposures are subdivided 2 or 4 times, and only begin to show a drop in persistence when the exposure is subdivided enough to avoid saturation.

## 3.3 Equivalence of electrical and optical stimuli

Non-destructive readout after a step increase in detector bias reveals an excess dark current which takes hours to decay away. This step increase in detector bias is the electrical equivalent of resetting and reading after a fixed photo-generated charge has been left on the diode for a very long period. The voltage drift observed is the equivalent of image persistence.

Responses to equivalent electrical and optical stimuli at four different amplitudes are shown in Fig. 6. When a light is pulsed at the start of the stimulus exposure it produces a persistence curve, which is very similar in shape to that generated by an electrically created voltage step on the diode of the same amplitude (induced by manipulating the reset level). The response to the electrical stimulus is typically measured to be 20% greater than the optical, though it is not yet clear whether this real or an artifact of our method.

Fig. 7 shows the same curves normalized by their final value, revealing that the initial response to an electrical stimulus is consistently slightly faster. We will present data below which shows that the response time constants depend on the stimulus timing which suggests again that experimental method may be responsible for this difference.

When the polarity of the voltage step is reversed to mimic a photo-generated signal (detector bias is stepped down, not up) then a signal droop is seen as the traps capture charge. This voltage change is smaller than that associated with an opposite voltage step which mimics resetting an optical signal (Fig. 6). If voltage changes occur in response to the motion of charges, and if trapping merely fixes the charge in place, why then is there any voltage change at all during the charge capture process? This phenomenon is of some interest, since it may account for the reports that sensitivity of HgCdTe detectors is flux dependent [2] [4], but we have not explored this further.
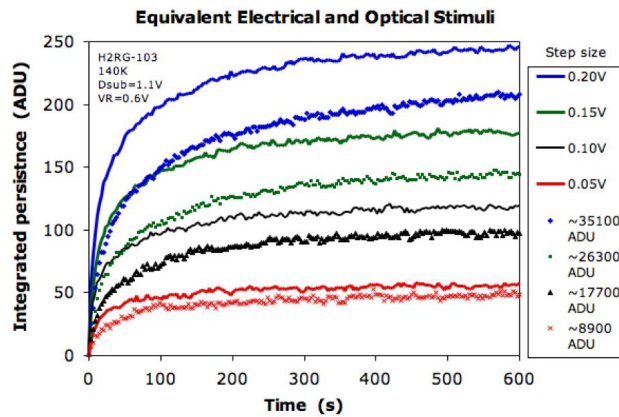


Fig. 6: Non-destructive readout after a 300s stimulus produced either by a light flashed at the start of exposure, or a step in the reset voltage. These produce strikingly similar results although the response to optical stimuli (symbols) is typically 19% lower than to the electrical equivalent (lines).
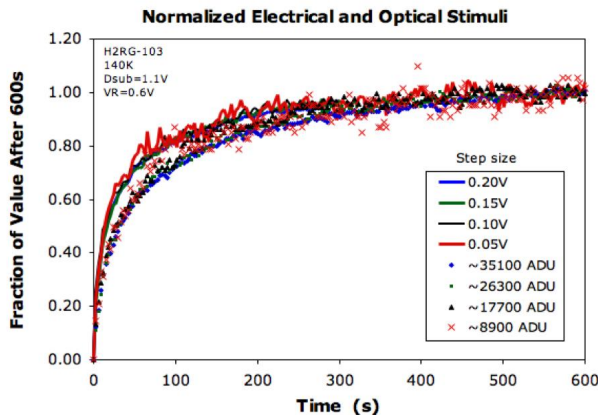


Fig. 7: Normalizing all curves in Fig. 6 to the same final value reveals that charge emission is consistently faster for the electrical stimulus, however the most striking feature is the overall similarity in the curve shapes.

### 3.4 Lower gain for persistence than for photo-generated charge

The total noise due to dark current, read noise, and persistence is measured by the spatial standard deviation of the difference of dark frames with similar position in the stimulus-response cycle. As the persistence decays, the noise

drops to that expected for dark frames much sooner than would be observed if photo-generated charge had accounted for the signal. Fig. 8 compares the variance of the persistence frames with the straight line observed for photo-generated charge.

If the noise obeys Poisson statistics, the slope of the variance-mean relation in Fig. 8 is equal to the number of data numbers (ADU) per electron. The clue to the correct explanation for the low variance is the increase in slope with mean, which indicates that noise is not sub-Poissonian but that gain (volts-per-persistence-electron) is increasing with stimulus amplitude.
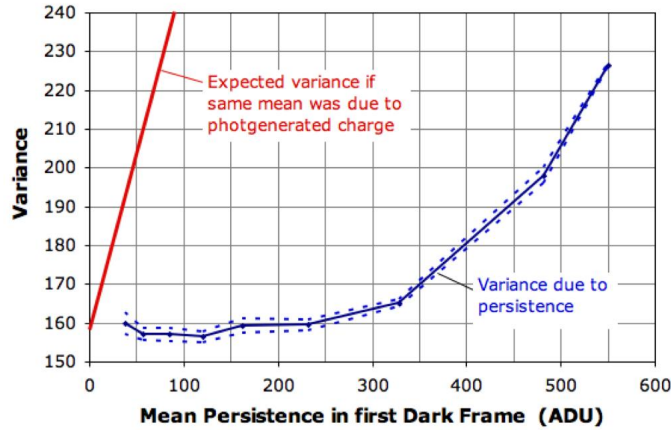


Fig. 8: The temporal variance measured for persistence is very low and slope increases with mean persistence. Contrast this with the linear variance-mean relation observed for photo-generated charge which has a measured slope of 0.93 ADU/e-. Dotted lines represent 1-sigma error in the estimate of variance.

The energy imparted to a charge by an electric field is the line integral of the electric field over the path taken by the charge. Voltage between two points is in turn the energy per unit charge to move the charge between those points. It follows that the voltage change on the diode junction should increase with the distance charge travels within the electric field in the depletion region.
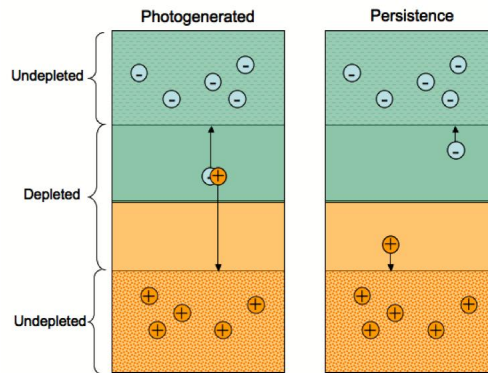


Fig. 9: Photo-generated electron-hole pairs travel a total distance equal to the full depletion width no matter where the photon is absorbed (left) so that photo-generated charge has constant voltage gain, at least until depletion width and electric field are modified by signal accumulation. Charge emitted by traps only travels a short distance back to the edge of the depletion region (right) so that gain (the slope of the variance-mean relation) for persistence is lower and highly dependent on stimulus amplitude.

The sum of the distances traveled by a photo-generated electron and corresponding hole is equal to the depletion width no matter where in the depletion region the electron-hole pair is generated (Fig. 9, Fig. 10). By contrast charge emission by traps occurs independently for electrons and holes (Fig. 9, Fig. 11). The distance traveled by de-trapped charge will be much less since traps are only filled near the depletion edge. The slope of the variance curve increases with brighter stimuli as traps further from the depletion edge become populated (Fig. 12). Even when the stimulus is bright enough for traps to be filled throughout most of the depletion region, traps close to the depletion edge are more likely to be filled

since they are exposed to charge for most of the stimulus exposure time: the mean distance traveled will be less than one quarter the depletion width so that variance for signal generated by persistence is always substantially lower than variance for photo-generated charge.
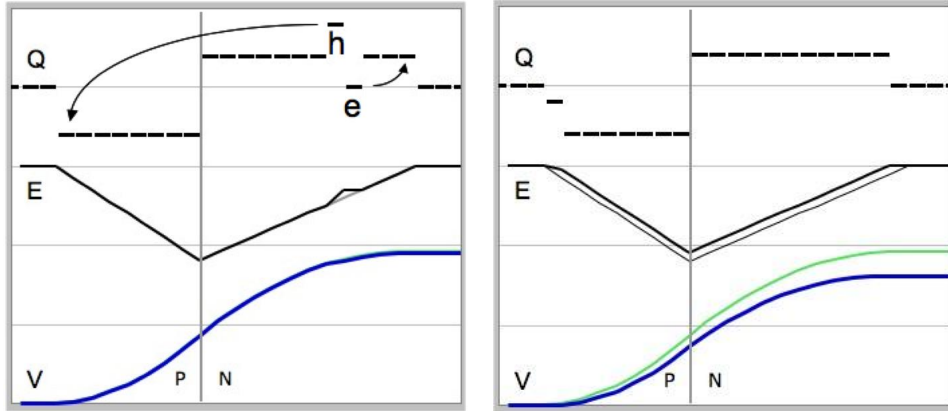


Fig. 10: The creation of the electron hole pair by the absorption of a photon does not immediately change the voltage across the diode (left panel) but the motion of the electric charge under the influence of the electric field does. The voltage change depends on the sum of the distances traversed through the electric field by the electron and the hole. This is equal to the depletion width and is not dependent on the location of the electron-hole pair creation.
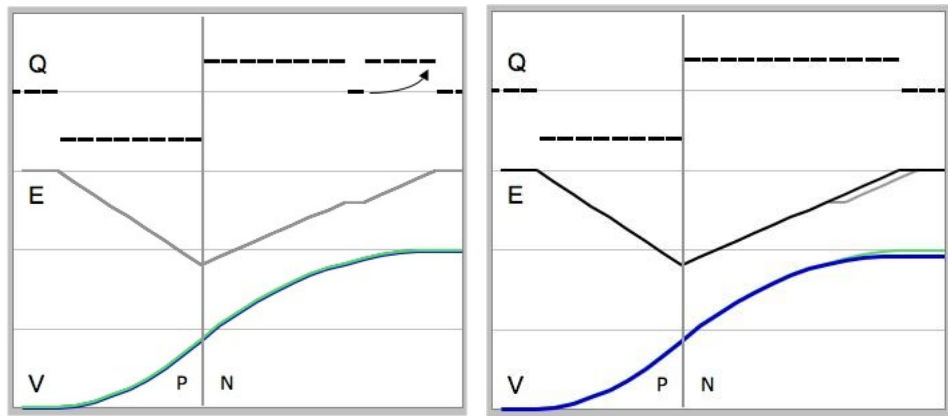


Fig. 11: An electron emitted from a trap near the depletion edge (left panel) produces relatively little voltage change (right panel) since it moves only a short distance through the electric field.
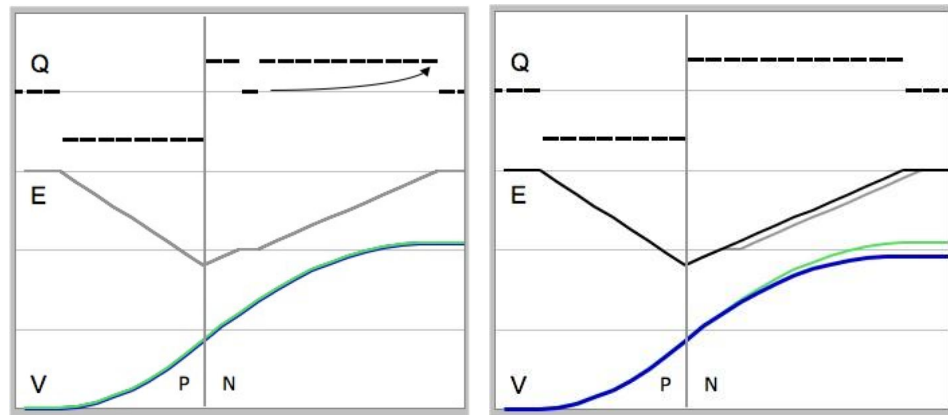


Fig. 12: An electron emitted further from the depletion edge (left panel) than in Fig. 11 produces a larger voltage change but never as much as a photo-generated electron which traverses the full depletion width (Fig. 10).

# 4.  PERSISTENCE LINEARITY, A PARADOX

Fig. 4 shows the persistence response for stimulus amplitudes reaching deep saturation.  Just prior to saturation, the persistence increases faster than linearly as would be expected given the increasing gain for charge emitted by traps further from the depletion edge.  However, over much of the normal signal range (where the response to photo-generated charge is fairly linear), the persistence is a surprisingly linear function of stimulus amplitude. At low stimuli the deviation from constant of the ratio of response to stimulus can be modeled by an additive error, which is equivalent to an excess persistence of just 0.75 electrons.
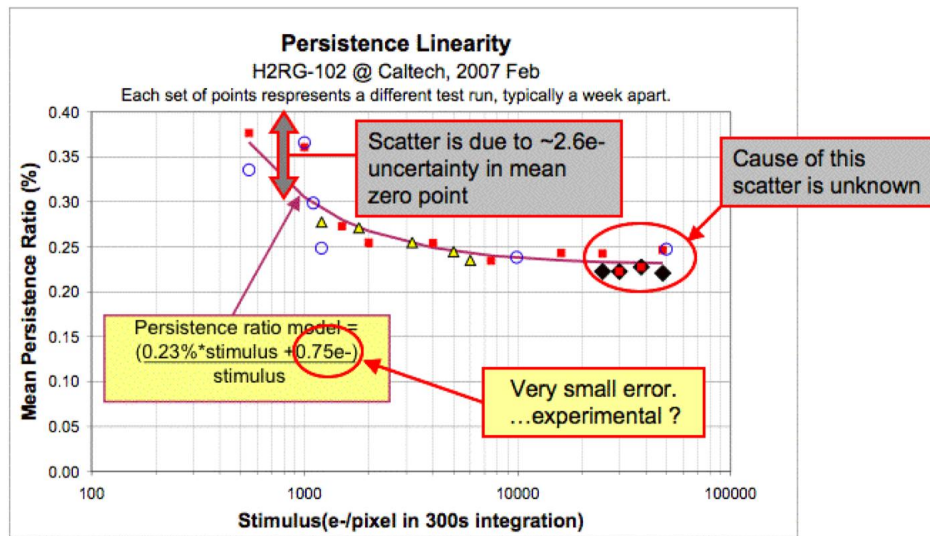


Fig. 13:   The persistence increases linearly with stimulus except for a slight (0.75e-) additive term which is unphysical (since it implies finite persistence when there is no stimulus) and thus likely to be due to experimental error.  The various symbols represent measurement runs conducted about a week apart showing long term stability better than +-0.02% of stimulus.  The cause of this scatter at high stimulus amplitude is unknown while the scatter at low amplitude is due to the 2.6e- rms drift in mean zero point for dark frames.  Detector bias was 0.5V. Reference pixel subtraction (per channel) and dark frame subtraction were employed.

The dependence of persistence gain on stimulus amplitude presents a paradox. How can persistence be linear versus signal (Fig. 4, Fig. 13) when gain inferred from the variance (Fig. 8) increases with distance from the depletion edge and thus with the stimulus amplitude?   The key to resolving this paradox may lie in examining the effect of doping and trap density profiles.  We do not have access to this information, but instead plan to run some simple numerical simulations for plausible density profiles.

# 5.  LOWERING PERSISTENCE BY DESIGN

## 5.1  Possible diode improvements

Fig. 14 shows schematically how the photodiodes are created on the surface of a uniform HgCdTe substrate by localized implantation of dopant atoms.   The resulting PN junction is bowl-shaped with its edge at the surface in the area surrounding the bump bond which provides contact to the multiplexor.   As signal accumulates the edge of the depletion region traverses both bulk traps and those lying at the surface.  Crystal dislocations and impurities which account for traps are almost certainly much more numerous at the surface than in the bulk material, so reduction in persistence is likely to require either better surface treatment to minimize trap density or some as yet unknown method to make a diode with the PN junction fully submerged in the bulk without reaching the surface (analogous to a buried channel MOSFET

or buried channel CCD).  Alternatively a way has to be found to reduce the movement of the depletion edge as the signal accumulates.  Higher doping density near the surface will do this but may have undesirable side effects.
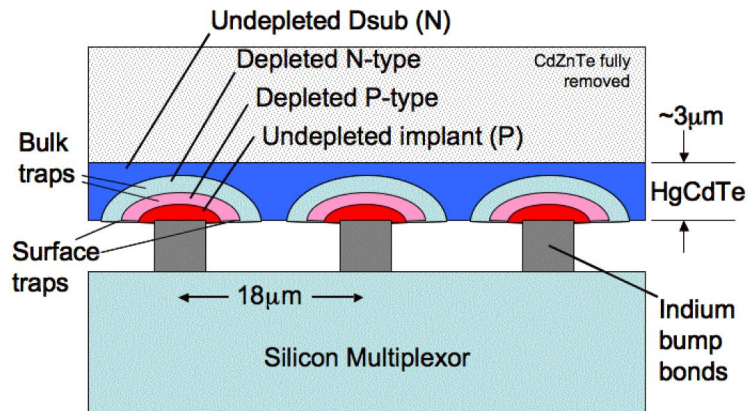


Fig. 14:  Schematic layout of HgCdTe photodiode array formed by implanting P type dopant into N type HgCdTe grown on a CdZnTe substrate which is removed after Indium bump bonding to a Silicon multiplexor.  The depletion region forms around the implant and thus curves up to surface where it is likely to encounter a much greater trap density than in the bulk material.  The level of persistence is thus likely to be a function of the surface treatment which will affect trap density and the doping density which will govern how much the depletion width will shrink as signal accumulates and thus the number of traps encountered.

## 5.2  Charge Trans-Impedance Amplifier

Persistence is a consequence of change in depletion width, which is inherent to integration of the signal on the capacitance of the isolated photodiode.  Instead of isolating the photodiode during the exposure, the diode can be connected to the summing node of a feedback circuit so that photo-generated charge is accumulated on the capacitance in the feedback path.  In this arrangement, called a Charge Trans-Impedance Amplifier (CTIA) or simply an "integrator", the voltage across the photodiode and thus the width of the depletion region remains constant and no image persistence should be observed.  The problem with the CTIA configuration is that it can be difficult to fit the integrator circuit into the footprint of a small pixel, and the circuit must be powered throughout the exposure instead of only when reading the signal.  Apart from the heat this produces, the photons emitted by the transistors in the CTIA have to be prevented from reaching the detector.

# 6.  MITIGATION

## 6.1  Exposure subdivision

Exposure subdivision (Fig. 5) may be used as a persistence mitigation scheme in special circumstances where there is sufficient signal for shot noise to dominate read noise, even when exposures have been subdivided.  For example it may be better to build up flat fields from a series of short exposures reaching rather a few longer exposures approaching full well.  e.g. If the mean signal is reduced from 50,000 e- to 1000 e- then the persistence in the first dark will be reduced from 115 e- to 1e- according the $N^{-1.22}$ scaling law in Fig. 5.

Wavelength calibration in a spectrograph typically relies on emission line sources with a mixture of very bright and faint emission lines.  If exposure time is set to overcome read noise in the faint lines, then the bright lines may produce considerable image persistence.  It is possible in the HAWAII-2RG to reset pixels selectively so that those pixels with plenty of signal could be exposed briefly many times while those, which are read noise limited would be exposed for longer, providing good S/N everywhere while minimizing persistence, at the expense of a custom readout waveform.

## 6.2 Time delay after reset

Fig. 15 shows the non-destructive readout of persistence accumulating after a 300s flat field exposure. Inserting a time delay between the initial reset and the first will read allow time for the faster traps to emit their charge before the delayed exposure start so that the perceived persistence signal will be reduced. This time delay reduces observing efficiency. For example, using the double exponential fit to the charge emission curve one can calculate that for a 300s exposure time with 30 sec readout time (the SNAP cadence), a 50% reduction in persistence comes at the expense of a 10% loss in observing efficiency. Further reduction in efficiency becomes rapidly more expensive since the longer emission time constant begins to dominate. This is unlikely to be a popular solution so we have turned our attention to calibration of persistence[1].
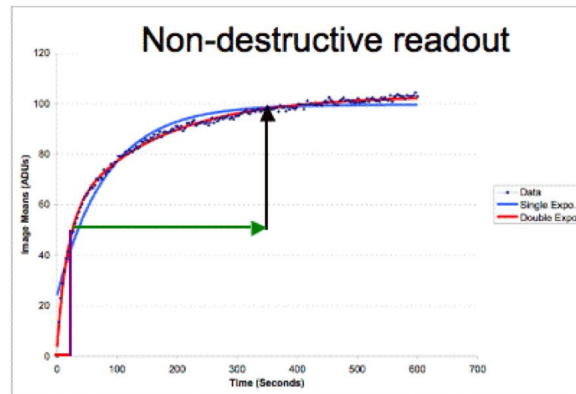


Fig. 15: A double exponential fit to the continuous non-destructive readout in a dark frame after a bright 300s exposure places equal power in time constants at 17s and 152s. As a consequence, delaying the exposure start by 30 sec will halve the persistence detected since the faster traps have time to emit charge before the start of the subsequent exposure. Of course this comes at the cost of lost of observing efficiency and is unlikely to be a popular solution.

## 6.3 Night light

We speculate that trapped charge resides at intermediate energy levels between the normal valence and conduction bands and that the slow emission of that charge is related to the low probability of a phonon (thermal vibration) providing the required energy to return the trapped charge to the conduction band when the detector is operated cold enough so that the same mechanism doesn't produce dark current by liberating electrons in the normal valence band. If this interpretation of the charge trap behavior is correct then it may be possible to accelerate charge detrapping by flooding the detector with photons beyond the cutoff wavelength such that traps are continuously being emptied without providing sufficient energy to raise normal valence electrons to the conduction band.

## 6.4 Flashy reset

A possible problem with the "night light" will arise if there are traps requiring energy close to (or greater than) the band gap, since photons at these wavelengths will produce a signal by exciting normal valence band electrons. However, *during reset* any photo-generated charge will be rapidly removed from the diode and the depletion width will be held constant, so it should be possible to flood the photodiode with photons both above and below the cutoff wavelength (i.e. above and below the band-gap energy). For this to be useful we require that traps be emptied rapidly enough that the observing time penalty is low since we can only conceal broadband illumination while reset is applied. This technique could be combined with the "night light" if necessary.

# 7. FURTHER WORK

Our first priority is to test the "Night Light" and "Flashy Reset" ideas to see if they will empty the traps efficiently and quickly enough to be useful, and to see whether there are adverse side effects such as filling new traps. A variety of illumination wavelengths will be tried and further insight into trap behavior may be gained by looking at emission time constants as a function of temperature. The observation that the persistence decay rate depends on exposure time leads us to hypothesize that there is a wide distribution of trap time constants and that we may require a wide range of photon energies to stimulate charge emission.

A significant theoretical challenge remains to explain the linearity of the persistence. Modeling hypothetical dopant and trap density profiles should provide some insight and these hypotheses may be tested by varying the detector bias (and thus initial depletion width

Until a solution to persistence is found we must continue to focus on learning how to calibrate this effect, so this is the topic of a companion paper[1].

# 8. ACKNOWLEDGMENTS

# 9. REFERENCES

[1] Smith, R.M., Zavodny, M, Rahmer, G. and Bonati, M., "Calibration of image persistence in HgCdTe photodiodes" Proc.SPIE 7021, paper 23 (2008)

[2] Aldering, G., et al., "Supernova / Acceleration Probe: A Satellite Experiment to Study the Nature of the Dark Energy", astro-ph, 0405232 (2004)

[3] de Jong, R. S., Bergeron, E., Riess, A. and Bohlin, R. C., "NICMOS count-rate dependent non-linearity tests using flatfield lamps" STScI Instrument Science Report NICMOS, 2006, report 001 (2006)

[4] Bohlin, R. C., Riess, A. and de Jong, R., "NICMOS Count Rate Dependent Non-Linearity in G096 and G141" STScI Instrument Science Report NICMOS, 2006, report 002 (2006)