

## Chapter 7

# Persistence in HyViSI Detectors

This chapter addresses one of the most prominent effects observed in HyViSI sensors: image persistence. Persistence in the HyViSI presents itself in a rich and complex way, as the following sections will reveal. But before delving into the observations, it is necessary to first present the theory and mathematics behind charge traps, which are commonly used to explain “persistent” effects in semiconductors. After this framework has been established, the data will be presented and an attempt to explain it with a somewhat complicated model will be made.

### 7.1 Trap Theory

Charge traps are one of the many nuisances in semiconductor optical detectors (and all semiconductors for that matter) and have been extensively studied for several decades. In CCDs charge trapping can degrade the charge transfer efficiency of the detector and render entire rows useless. In Hybrid CMOS detectors, we find that the traps give rise to latent images, signal decays after cosmic ray events, and high dark current after forward biasing the photodiodes. Further, the emission and capture of charge carriers by the traps may be one of the most important mechanisms for generating cross-talk between pixels.

A *trap* is actually nothing more than a *deep impurity* in the semiconductor crystal. These impurities have states which lie in the band gap far away from the conduction and valence band edges, typically requiring an energy of  $\sim 5kT$  to ionize. They are thus very efficient centers for recombination in the crystal. The term *deep* distinguishes them from *shallow* impurities, which are the kind of impurities that are purposely implanted in the semiconductor in the doping process. *Shallow* impurities typically lie at about  $\sim kT$  away from the band edges, so a large fraction of them are ionized purely by thermal energy.

If we were to look at a silicon crystal doped with impurities at the subatomic level, we would find that traps are constantly being filled and vacated. The idea is somewhat abstract, especially

since holes are fictitious particles. But we imagine the electrons and holes are trapped in energy space somewhere in the bandgap (below  $E_c$  and above  $E_v$ ) and that these traps also correspond to physical locations in the lattice. Electrons are captured at a rate  $c_n$  and emitted at  $e_n$  while holes are being captured at  $c_p$  and emitted at  $e_p$ . At equilibrium, all these processes are balanced so that the number densities of holes in the valence band,  $p_v$ , and electrons in the conduction band,  $n_c$ , remain essentially constant when measured by a macroscopic means, i.e.  $dp_v/dt = dn_c/dt = 0$ .

The processes of capture and emission for holes and electrons can be summarized as follows (with the corresponding coefficient listed):

- **Electron Capture** -  $c_n$  : An electron drops from the conduction band into a trap.
- **Electron Emission** -  $e_n$  : An electron jumps to the conduction band from a trap.
- **Hole Capture** -  $c_p$  : An electron in a trap drops into a hole in the valence band.
- **Hole Emission** -  $e_p$  : An electron in the valence band jumps into a trap and leaves a hole in the valence band.

The processes of *generation* and *recombination* of carriers can occur with combinations of these. Janesick provides two useful statements [3] to help understand how these processes occur **1) Electron and hole capture are required for recombination** and **2) Hole and electron emissions must occur nearly simultaneously through the trap in order to generate a dark carrier.**

If the temperature of the semiconductor crystal is changed rapidly or light is shined on it, the system will be brought away from equilibrium. The non-equilibrium situation was examined in detail by Shockley and Read [112], who formulated the mathematics that describe the role of traps in recombination and generation in a semiconductor. Among the numerous applications of their theory, McNutt and Meyer [113] used the statistics of traps and their temperature dependence to identify impurities in buried channel CCDs and Solomon [42] used Shockley-Read traps to model persistence in InSb hybrid infrared detectors. The reader is referred to these references for a detailed discussion of trapping statistics.

For this discussion, the notation in [113] will be adopted.  $N_{TT}$  will denote the total bulk state concentration of traps,  $n_T$  will denote the concentration of those traps filled with electrons, and  $p_T$  will denote the concentration of those traps filled with holes. It follows that

$$N_{TT} = n_T + p_T. \quad (7.1)$$

The dynamics of the free carriers in the system is governed by

$$\frac{dn}{dt} = \frac{dn_c}{dt} = e_n n_T - c_n n p_T \quad (7.2)$$

$$\frac{dp}{dt} = \frac{dp_v}{dt} = e_p p_T - c_p p n_T, \quad (7.3)$$

where  $n = n_c$  is the concentration of electrons in the conduction band and  $p = p_v$  is the concentration of holes in the valence band. The subscripts in  $n_c$  and  $p_v$  will be left off for notational convenience. In a conserved system where no free carriers are introduced from outside, trapped electrons can either be stolen from free electrons in the conduction band or gained through the donation of a free hole to the valence band. A similar argument for trapped holes leads to the equations:

$$\frac{dn_T}{dt} = c_n np_T - e_n n_T - c_p pn_T + e_p p_T \quad (7.4)$$

$$\frac{dp_T}{dt} = -c_n np_T + e_n n_T + c_p pn_T - e_p p_T \quad (7.5)$$

It is worth noting that the number of conserved systems in the detector is very limited. In fact, the only system that will be treated in this way is the *entire detector layer* while it is not exposed to light and the pixels are integrating dark current. The number of free holes introduced through leakage currents and electrons lost to the external load supplying  $V_{SUB}$  will be negligible. Carriers emitted by traps become free carriers and may relocate inside the detector, but they are assumed not to escape.

In a non-conserved system, the equations for  $dp/dt$  and  $dn/dt$  have additional terms. Take the  $p^+$  implant in the detector pixel, for instance. The number of free holes in the implant increases with time ( $dp/dt > 0$ ) as they are introduced by photogenerated carriers from the depletion region or surface/bulk thermal leakage currents. At reset, electrons are injected into it through the reset transistor channel in the multiplexer to recombine with the free holes ( $dp/dt < 0$ ), which brings its potential to  $V_{RESET}$ . If traps outside the implant emit holes, these holes may be added to the implant as free carriers as well ( $dp/dt > 0$ ), and the migration of these holes will cause a rise in signal. In this last case, even though the sum of free carriers and trapped charge in the detector may be conserved, their redistribution in space can result in an increased electric potential at the collecting node of the pixel. In fact, Solomon attributed latent images in InSb arrays to holes being emitted from traps outside of the  $p^+$  implant and collected inside of it [42]. The detector was considered a conserved system in which the holes emitted by traps translated directly into free carriers that were swept into the  $p^+$  implant and counted as signal. The implications of this model will now be examined and later it will be shown that it cannot account for persistence in HyViSI devices.

### 7.1.1 Hole Capture and Release from Shallow Traps

Solomon attributed persistence in InSb hybrid CMOS detectors entirely to holes being trapped in the surface interface between the bulk InSb and  $\text{SiO}_2$  passivant during illumination (see Figure 1.7). When the holes are emitted from the traps in subsequent exposures, they either drift or diffuse to the collecting node and cause a rise in signal. Thus, persistence is described entirely through the  $c_p$  term in Equation 7.5, which accounts for holes being captured during illumination, and the  $e_p$  term

in Equation 7.3, which accounts for holes being emitted in subsequent exposures and collected as signal charge.

According to this model, if the detector pixels are illuminated with sufficient flux or fluence and holes are introduced at the rate  $p_\gamma$ , then at steady state

$$\frac{e_p}{c_p} = p_\gamma e^{F-E_T/kT}, \quad (7.6)$$

where  $F$  is the Fermi Level (or quasi-Fermi level in the non-equilibrium case) and  $E_T$  is the trap energy. If  $E_T$  lies just above the valence band then the capturing processes will dominate emission and there will be a surplus of trapped holes relative to when the detector was not illuminated. The capture cannot continue indefinitely, of course. It goes like  $p_\gamma c_p n_T$  and cuts off when the number of electron-filled traps,  $n_T$ , goes to zero. When the light source is removed or the array has been shuttered, the rates will change according to

$$\frac{e_p}{c_p} = N_V e^{E_V-E_T/kT}, \quad (7.7)$$

where  $N_V$  is the effective density of states in the valence band and  $E_V$  is the valence band energy. The trapped holes will thus be re-emitted in a subsequent exposure with a time constant of  $e_p^{-1}$  and generate a signal, even in the absence of illumination. And since the traps are *shallow* (they are located much closer to  $E_V$  than  $F$ ), the emission occurs at a very quick rate en route to equilibrium. These emitted holes are assumed to be the source of latent images.

The signal in a subsequent exposure will depend on the excess of holes trapped during illumination,  $\Delta p_T$ , the time since the source of illumination was removed,  $t_o$ , and the integration time of the exposure, i.e. the time since reset,  $\Delta t$ . From the above equations, we can derive the following formula for the signal imparted by the release of holes,  $S(t_o, \Delta t)$ :

$$S(t_o, \Delta t) = \Delta p_T e^{-t_o/\tau} (1 - e^{-\Delta t/\tau}), \quad (7.8)$$

where  $\tau = e_p^{-1}$  is the time constant associated with the release of the holes from the traps. This expression represents the case in which there is only one trap species with one time constant. However, we might also imagine that a whole slough of impurities are present, and that each has its own associated time constant  $\tau_i$  and number of traps  $N_{T_i}$ . Equation 7.8 then becomes a linear combination over those species:

$$S(t_o, \Delta t) = \sum_i^N \Delta p_i e^{-t_o/\tau_i} (1 - e^{-\Delta t/\tau_i}), \quad (7.9)$$

If we differentiate this equation with respect to the integration time, we find that the rate of change

of the signal (the instantaneous dark current,  $D$ ) goes as

$$D(T) = \frac{dS(t_o, \Delta t)}{d\Delta t} = \sum_i^N \frac{\Delta p_i}{\tau_i} e^{-(t_o + \Delta t)/\tau_i}. \quad (7.10)$$

We see that the total time,  $T = t_o + \Delta t$ , is what actually determines the rate of signal change in any subsequent exposure. And further, this rate should decrease monotonically with time after the source has been removed. One way of interpreting this is that the traps are being depopulated regardless of whether the detector is set to integrate or is being held in reset. We can measure  $\tau$  by performing integrations and fitting  $S(T)$  with the function in Equation 7.8, or some linear combination of these functions which represent different trap species, each having their own density and time constants. The expressions tell us that if we wait a time  $t > 3\tau_l$ , where  $\tau_l$  is the longest time constant involved,  $dS/dt$  should be essentially zero. Another way of saying this is that if we wait long enough — whether we integrate, reset, or leave the detector idle while waiting — the trapped holes should all be released and not interfere with the signal in the next exposure. Once  $dS(t_o, \Delta t)/d\Delta t \rightarrow 0$ , it should remain at zero until we populate more traps with illumination. Since this **does not happen in the HyViSI detectors**, another model for the persistence needs to be introduced.

### 7.1.2 Hole and Electron Capture from Deep Level Traps

If the traps are deep traps<sup>1</sup> with energies  $E_T$  that lie close to the Fermi level  $F$ , the situation changes from the one just described. Thermal emission rates are proportional to a Boltzmann factor,  $\exp(-\Delta E/kT)$ , where  $\Delta E$  is the depth of the trap (free energy) from the band edge to which the carrier is emitted [114]. The electron and hole emissions will be governed by

$$e_n = A_n e^{-(E_C - E_T)/kT} \quad (7.11)$$

$$e_p = A_p e^{-(E_T - E_V)/kT}, \quad (7.12)$$

where  $A_n$  and  $A_p$  are coefficients related to the density of states at the conduction and valence band edges, respectively, as well as the capture cross sections (both coefficients go like  $T^2$ ). Thus, if  $E_T$  lies far from both  $E_C$  and  $E_V$  the emission rates should be small and capturing events should dominate.

In the case of the SiPIN detector, if the bulk material and surface interface between the bulk and SiO<sub>2</sub> passivant is fully depleted and in thermal equilibrium (in the dark), then  $n \approx p \approx 0$ . If deep level traps exist in these regions, they cannot capture carriers in an equilibrium state simply because

<sup>1</sup>As Miller et al. point out [114], the names applied to deep defect states—traps, recombination centers, generation centers, deep levels, deep impurities, and so on—can be quite confusing. For the argument presented here, it will imply that the trap energy lies only a few  $kT$  away from the Fermi level.

carriers are not available. But with illumination, reset, or forward bias, both of these regions see an increase in free carriers that are available for trapping. As in Solomon’s model, under illumination the number of hole filled traps increases at the rate  $p_\gamma c_p n_T$ , and when the array is finally shuttered there will be an excess of hole filled traps  $\Delta p_T$ . However, because the traps are deep in the forbidden energy gap, the holes will not be efficiently emitted from the traps. Instead, they will be “stuck” at whatever atom plays host to the trap until an electron is available to return the trap to its initial charge state.

The electrons necessary to repopulate the hole filled traps can be provided by closing the reset switch. The function of the reset is to provide the carriers necessary to restore the potential at the integrating node to  $V_{RESET}$ . Ideally, the reset would provide an ample amount of electrons in a given pixel  $\Delta n_T$  to repopulate the traps in that pixel at the rate  $n c_n p_T$ . But if some of these electrons migrate to other regions of the detector instead of being captured by the traps, the full restoration will not happen. Electrons that quickly escape and drift all the way to the back surface of the detector will cause a dramatic rise in the potential<sup>2</sup> and those that diffuse horizontally may cause an decrease in signal for the neighboring pixels. Once all of the electrons in the pixel have either recombined or left the collecting node and a quasi-equilibrium state has been reached,  $dS(t_o, \Delta t)/d\Delta t \rightarrow 0$ . However, the fraction of hole filled traps that remain at the time of the next reset will cause the same effect, giving rise to a nonzero value for  $dS(t_o, \Delta t)/d\Delta t \rightarrow 0$  in the following exposure, but with a diminished level. This is the crucial difference between the shallow trap model and the deep trap model, one that explains how the persistence can be “recharged” after a reset. In Section 7.5, the deep trap model will be considered in more depth to account for other features of the persistence in HyViSI detectors.

---

<sup>2</sup>Actually, applying Poisson’s equation and integrating the electric field shows that even the electrons that simply migrate a small distance to the trapping sites will also cause an increase in the potential. But because the electrons that drift all the way to the back surface travel further in the electric field, they generate a larger change in potential.

## 7.2 Latent Images and Persistence

Hybrid CMOS detectors are prone to an effect called "persistence" in which regions of pixels that were previously exposed to significant brightness seem to "re-emit" that brightness after they have been reset. The "re-emission of brightness" forms a *latent image* of the object that emitted the light. CCDs show a similar phenomenon referred to as Residual Bulk Image (RBI). Both phenomena are usually attributed to the release of charge from *traps* somewhere in the detector, [42, 24, 113] usually near the surface of the bulk material closest to the collection node.

In the HyViSI detectors, a different type of this "persistence" effect is seen. Figure 7.1 shows a very dramatic case of this in a raw frame taken from an exposure of Saturn. The white shape at top is a saturated image of Saturn and five of its moons. Below it is a latent image from the previous exposure taken before the telescope pointing was offset to move the light from Saturn across the detector.

What is unique is that we do not only see regions of positive brightness in the subsequent exposures; we also see dark pixels surrounding the bright ones. Of course, in this sense brightness and darkness are simply terms that relate the signal level to the background far away from the affected area since the persistence shows itself even in exposures taken when the detector is in the

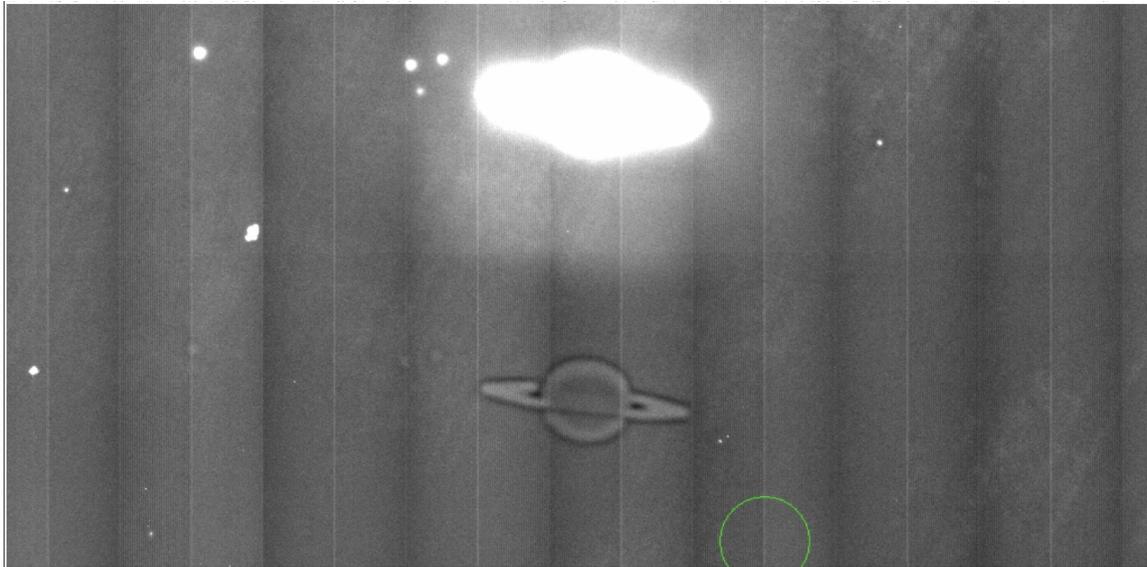


Figure 7.1: An extreme example of persistence in HyViSI detectors. This is a raw read (no subtraction or manipulation of pixel values) from an up the ramp exposure of Saturn. Five of its moons are visible: Titan, Dione, Tethys, Enceladus, and Rhea (from left to right). It is tempting to wonder whether the persistence might serve a useful purpose since it provides contrast information from a previous, completely saturated exposure that contained none.

dark. Because there is no inversion of the signal taking place, the regions of brightness are due to pixels that see a rapid increase in voltage with time. And because the pixels are collecting minority carriers (holes generated in the n-type substrate), this indicates that holes are accumulating in the  $p^+$  implants of those pixels or electrons are leaving the implants. The regions of darkness are due to pixels that see a decrease in voltage with time, indicating that holes are leaving those  $p^+$  sites, or rather that they are being populated by electrons.

Latent images present themselves in two distinct and fundamentally different ways in the HyViSI detectors: **1) persistent charge emission** and **2) slow-decay offset** or **Laser Burn-In**. The first of these gives rise to ephemeral latent images that possess the aforementioned dark and bright regions and vanish on the order of minutes or hours. The after-image of Saturn in Figure 7.1 is an example of persistent charge emission. The second type can last for months or years. These “burned-in” images occur for very large flux, and are hypothesized to be due to stressing of the source follower gate oxide in the multiplexer unit cell. They will be discussed in Section 7.3.

## 7.2.1 Persistent Charge Emission

As mentioned in Section 7.1.1, numerous studies have used Shockley-Read trapping physics to explain persistence in hybrid detectors. Under these assumptions, the time constant associated with a given trap determines the rate at which the traps are depopulated, and consequently, the rate at which the pixel signal changes. The data collected in our experiments show that there are multiple time constants involved with latent image decays, and in fact, traps that seem to be empty can be recharged by a reset of the detector. This behavior has been noted by several authors in the study of changes in the flatband voltage of Si-SiO<sub>2</sub> MOS capacitors [115]. They designated these rechargeable centers as “Anomalous Positive Charge” centers (APC), and showed they had a duration on the order of 60 minutes. HyViSI detectors exhibit a similar type of “rechargeable”, long-term shift in operating voltages. However, there is evidence to indicate that the shift is due to diffusion of trapped charge carriers rather than an anomalous positive charge. In the next section, an introductory look at the signature of the latent images is given.

### 7.2.1.1 Spatio-Temporal Dependence of Persistence

Figures 7.2 and 7.3 are meant to convey the spatial and temporal behavior of persistent charge after saturation from a bright star. The first of these is a time sequenced set of 40 second up-the-ramp exposures taken with H2RG-32-147. In the very upper left plot of Figure 7.2, the radial profiles from sequential reads of the star in a saturating exposure are shown. The plots show the pixel values after the first read has been subtracted, so the fact that the radial profile close to the end of the exposure has a large divot near the center indicates those pixels have accumulated enough charge to rail the detector output, and presumably saturate the well. The image at the very upper right was taken from the last read of the same exposure and is simply a two dimensional representation

of the latter. In the other seven plots and images taken from the seven subsequent exposures, the telescope has been moved and the star is no longer in the picture. No star is present, so only the sky background light should generate signal in the image. However, the detector somehow “remembers” that a bright star was there previously, even though 15 resets were applied between each exposure. Its memory of the star shows itself as a distorted image that not only has a peak or “core”, but a surrounding valley or “halo”. The latent image deteriorates with time until after several minutes time, the detector “forgets” the star was there. It should be noted that all data shown in Figure 7.2 were taken with the detector exposed to the sky, so a non-negligible photon flux causes the signal to rise even after the persistence has subsided.

The spatial structure of the persistent image is especially interesting. Figure 7.3, a radial profile from a dark exposure taken immediately following a 35.4 second  $I$  band exposure, encapsulates some of the interesting features. Again, the numbers take negative values because the first read value has been subtracted from each pixel. The radial center  $r = 0$  in the figure coincides with the centroid of a saturated star in the  $I$  band exposure. At this radius, the persistence signal has its maximum value  $I_{max}$ . At a radius of  $R_{core}$ , the pixels have seen *no net change* in signal. They may have risen and fallen back or vice versa, but they have returned to the voltage at which they started. The radius  $R_{min}$  coincides with the location where the pixels have seen the greatest *net decrease*  $I_{min}$ . And further out,  $R_{Halo}$  is the radius at which the pixels have again seen no net change. Beyond this radius, the pixel signals have not yet been affected by the persistent “cloud”. Also shown in the plot is the positive core region shaded in blue and the negative halo in yellow.

The pixel values are reasonably well fit with the following empirical expression (shown in green):

$$I(r, t_0, \Delta t) = I_{max}(t_0, \Delta t) * \cos(\alpha(t_0, \Delta t)r^{\beta(t_0, \Delta t)})e^{-r^2/\gamma(t_0, \Delta t)} \quad (7.13)$$

where  $t_0$  is the time since the array was shuttered (or since the illuminating or *offending* source was removed),  $\Delta t$  is the time since the array was reset, and  $\alpha, \beta$ , and  $\gamma$  are, in general, increasing functions of  $t_0$  and  $\Delta t$ . It is important to note that no particular physical significance has been found in this expression; it is purely empirical. Moreover, the fit only works under certain limiting circumstances such as short exposure times, non-saturated pixels, and minimal PSF broadening due to atmospheric blur or tracking error. However, it has utility in some cases, as it may be used to fit and remove a persistent image from a subsequent exposure. It also provides insight into the structure of the persistent shape and how it evolves in time.

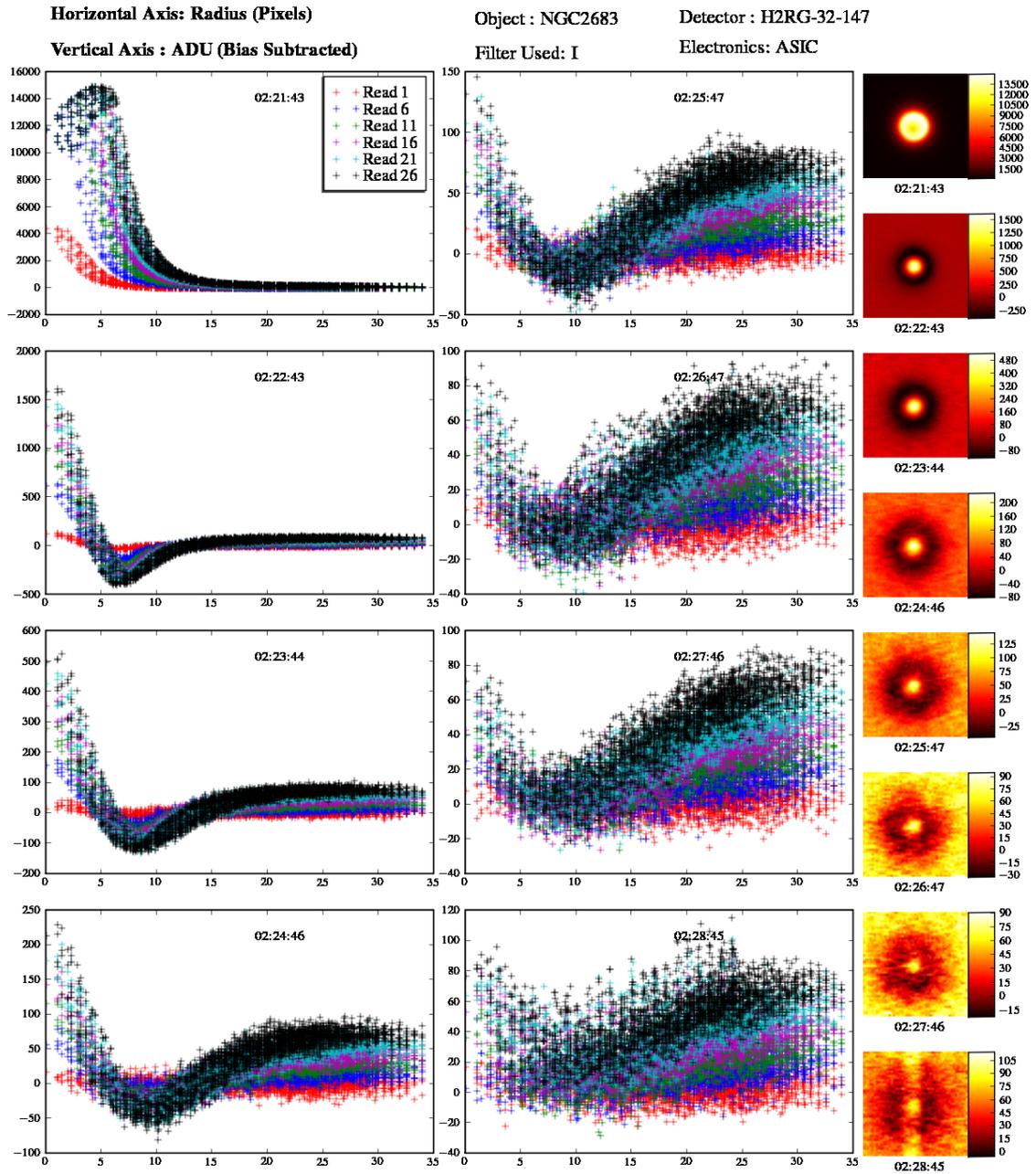


Figure 7.2: Radial profiles and image snippets taken from 30 read up the ramp exposures illustrating the evolution of persistence (starting at top left) in HyViSI pixels. The top left plot and upper image are of an actual star. In the other seven plots and images, no star is present, but persistent charge emission generates signal in the pixels.

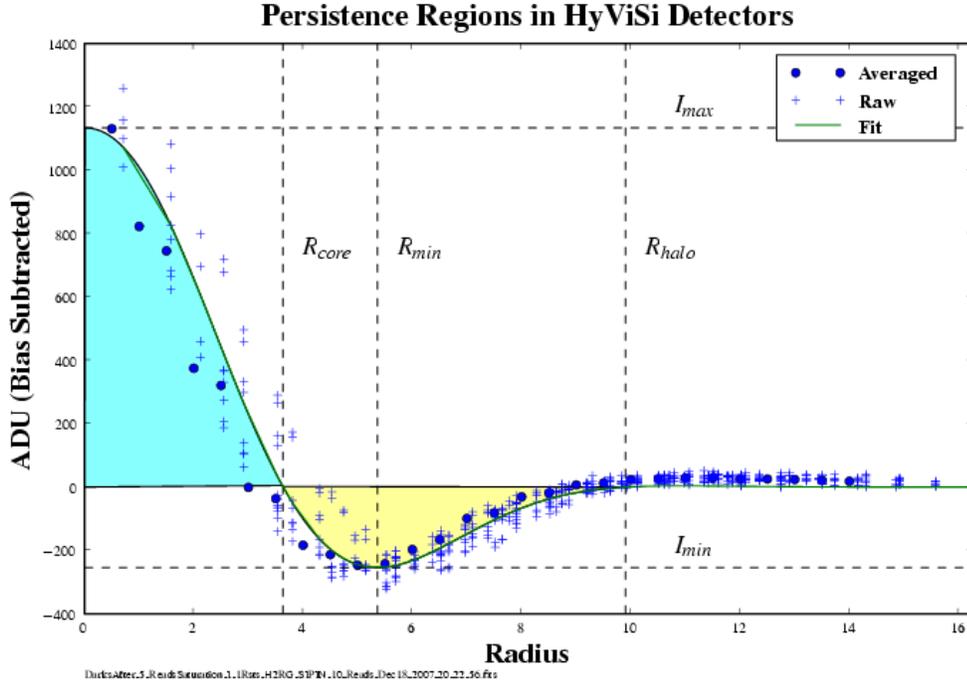


Figure 7.3: Radial profile for a region from a dark exposure. The radius is measured in pixels. A bright star was imaged in the same region in an  $I$  band exposure taken immediately preceding the dark. The blue shaded area represents the persistence core and the yellow shaded region represents the halo in this “core-halo” persistence. Several other coordinates are noted.

Although Equation 7.13 and the two figures shown do not address certain subtleties with the persistence (dependence on flux and fluence, wavelength dependence, etc), they do convey several key features.

1.  $I_{max}(t_0, \Delta t)$  is an increasing function of  $\Delta t$

In an up-the-ramp integration it has a preliminary exponential dependence on the exposure time as it rises to some maximum value  $I_{Full}$  and goes like

$$I_{max}(t_0, \Delta t) = I_{Full} * (1 - \exp^{-t/\tau_{rise}}), \quad (7.14)$$

where  $\tau_{rise}$  is the time constant for the rise.  $\tau_{rise}$  depends on temperature, biasing, operating conditions, etc., but is typically on the order of seconds or minutes. As shown in Figure 7.4, for very long exposures,  $I_{max}$  will tend to decrease after it has reached  $I_{Full}$  in a quasi-linear fashion. This last aspect makes it very difficult to fit latent images in long exposures.

2.  $I_{Full}(t_0, \Delta t)$  is a decreasing function of  $t_0$

It has an exponential dependence on  $t_0$  as it decays to zero:

$$I_{Full}(t_0, \Delta t) = I_{Full}(t_0 = 0) * (exp^{-t/\tau_{decay}}), \tag{7.15}$$

where  $\tau_{decay}$  is the time constant for the decay. In some cases, it is on the order minutes; in others, hours. The decay time has a dependence on the number of resets and reads performed since  $t_0 = 0$ .

**3.  $I_{Full}(t_0, \Delta t)$  is proportional to the incident flux/fluence in the offending exposure.**

It shows a linear relationship to the fluence in the stimulus image in cases where the detector pixels were at or below saturation in the offending exposure. Above saturation, the relationship becomes nonlinear and eventually  $I_{Full}$  reaches its own saturation threshold.

**4.  $\cos(\alpha(t_0, \Delta t)r^{\beta(t_0, \Delta t)})$  accounts for the oscillation from the positive core to the negative halo**

The fact that  $\alpha$  and  $\beta$  are both increasing functions of the two time variables means that the halo is moving outward with time. This also means that  $R_{Halo}$  increases faster than  $R_{Min}$  and  $R_{Min}$  increases faster than  $R_{Core}$ .

**5.  $e^{-r^2/\gamma(t_0, \Delta t)}$  indicates the structure is diffusing radially outward**

It effectively represents the fact that the overall core-halo structure is diminishing in amplitude and getting broader with time. The diffusive behavior of the structure suggests that its dynamics are, at least in part, governed by the diffusion of charge carriers.

Each of these 5 points, along with some subtleties in the core-halo behavior, such as the existence of multiple time constants, will be discussed in further detail in the sections that follow.

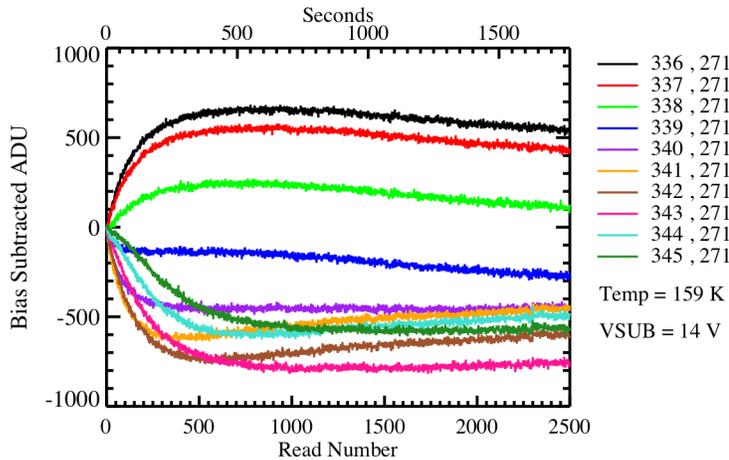


Figure 7.4: Persistence signal in a very long up the ramp exposure after saturation from a bright star. The signal vs. time is shown at different radii from the center of the core-halo image (pixel 336, 271). out toward the halo. Note that the pixels near the core initially see a rapid gain in signal followed by a slow, quasi-linear loss.

## 7.2.2 Dependence on Flux and Fluence

As Smith et al. note, it is a common misconception that persistence is caused by saturation or that it is a result of improper pixel resets [116]. For both SiPIN and per-pixel depleted infrared detectors, both of these statements are fallacious. In fact, persistence seems to be a process that is as intrinsic to the inner workings of a semiconductor detector as the photoelectric effect itself is. In this section, experimental results are shown that indicate latent images form well below the saturation threshold in HyViSI detectors. The dependence of the persistent signal on the flux and fluence of the *offending* stimulus will now be described.

### 7.2.2.1 Experimental Description

To probe a large range of brightnesses in a short time, the open clusters NGC 9256 and NGC 2395 were observed with H2RG-32-147 in full frame mode. The detector was operated at a temperature of 170 K and a substrate voltage of  $V_{SUB} = 15$  V. No idle resets were performed in between exposures. To measure persistence, the following procedure was applied:

- 1) An UTR exposure of the cluster with  $N_{Reads}$  reads was taken. Each read took approximately 1.41 seconds.
- 2) The filter wheel was moved to the blank position so the detector was no longer illuminated.
- 3) A set of 20 dark exposures was taken. Each dark exposure consisted of 1 reset and 10 reads. The read time was 1.41 seconds. The integration time for the 10 reads was 14.4 seconds and approximately 24 seconds elapsed (the IDL scripts took some time to write the file and header) between the start of consecutive exposures.

This process was repeated in each filter band ( $g$ ,  $i$ , and  $y$ ) for  $N_{Reads} = 5, 10, 15, \dots, 40$ . The range of brightnesses and exposure times provided an ample sampling over the detector well depth and beyond saturation.

### 7.2.2.2 Core Maximum and Halo Minimum

Figures 7.5 and 7.6 show the maximum and minimum persistence signal, respectively, attained in the first dark exposure of step 3) as a function of the maximum stimulus signal in step 1). Each data point represents a core (halo) value for a particular star in a particular image. For instance, a star that fills up 25% of the full well in the center pixel for  $N_{Reads} = 5$  will have a corresponding data point at about 50% of full well for  $N_{Reads} = 10$ , 75% for  $N_{Reads} = 15$ , etc. An estimate for stars that saturate the detector quickly in the stimulus images is made by extrapolating the photocurrent slope before saturation.

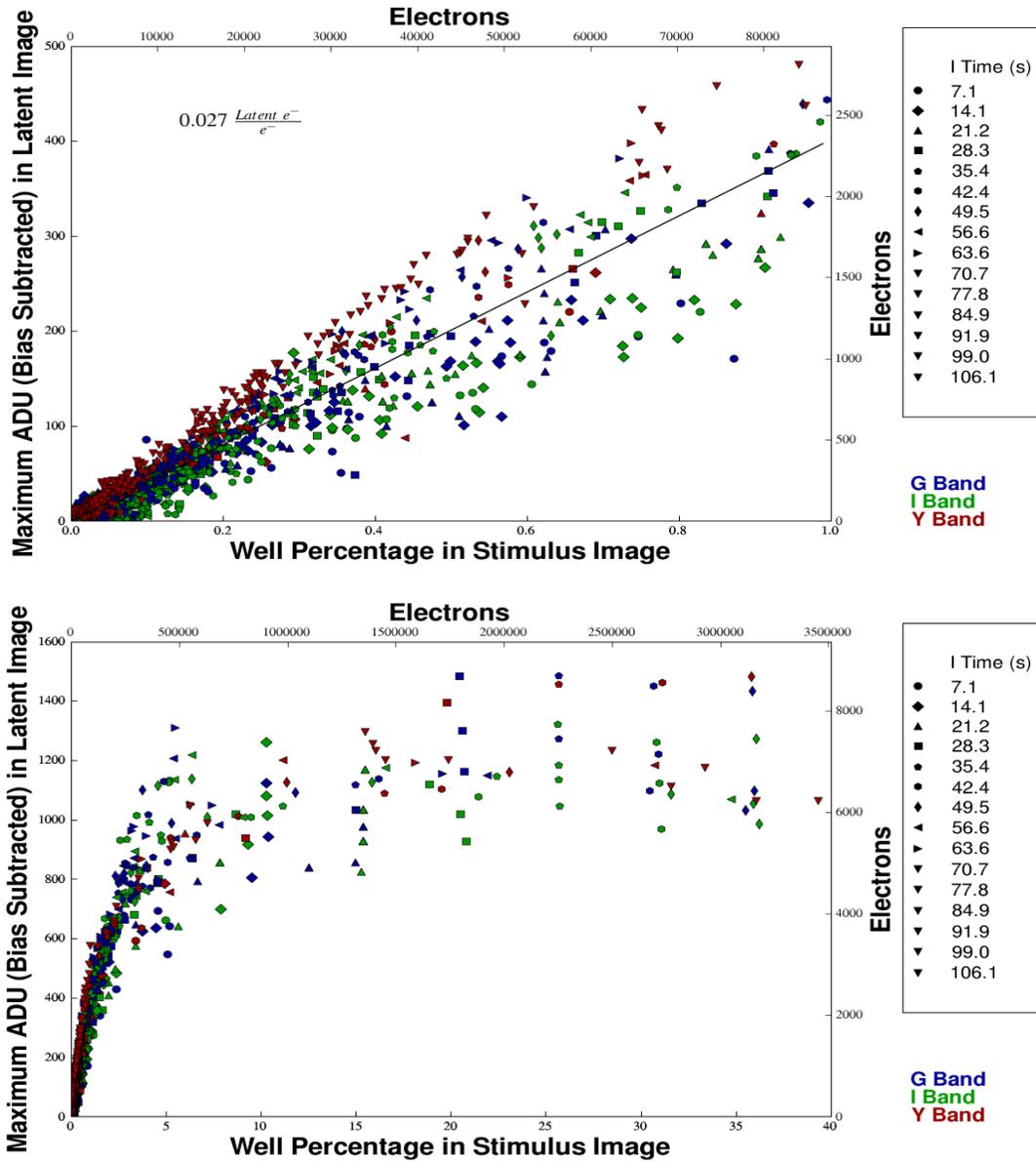


Figure 7.5: These plots show the relationship between persistence signal in the *core* and the stimulus that caused it. On the *x* axis is the maximum signal attained by the illuminated stimulus (fluence). Values beyond full well are extrapolated by fitting a slope to the pixel before it is saturated. On the *y* axis is the maximum persistence signal attained in a 14.4 second dark exposure following the illuminated stimulus. Each data point in these plots corresponds to a particular star in a particular image. The integration time for the point is listed on the legend at the right.

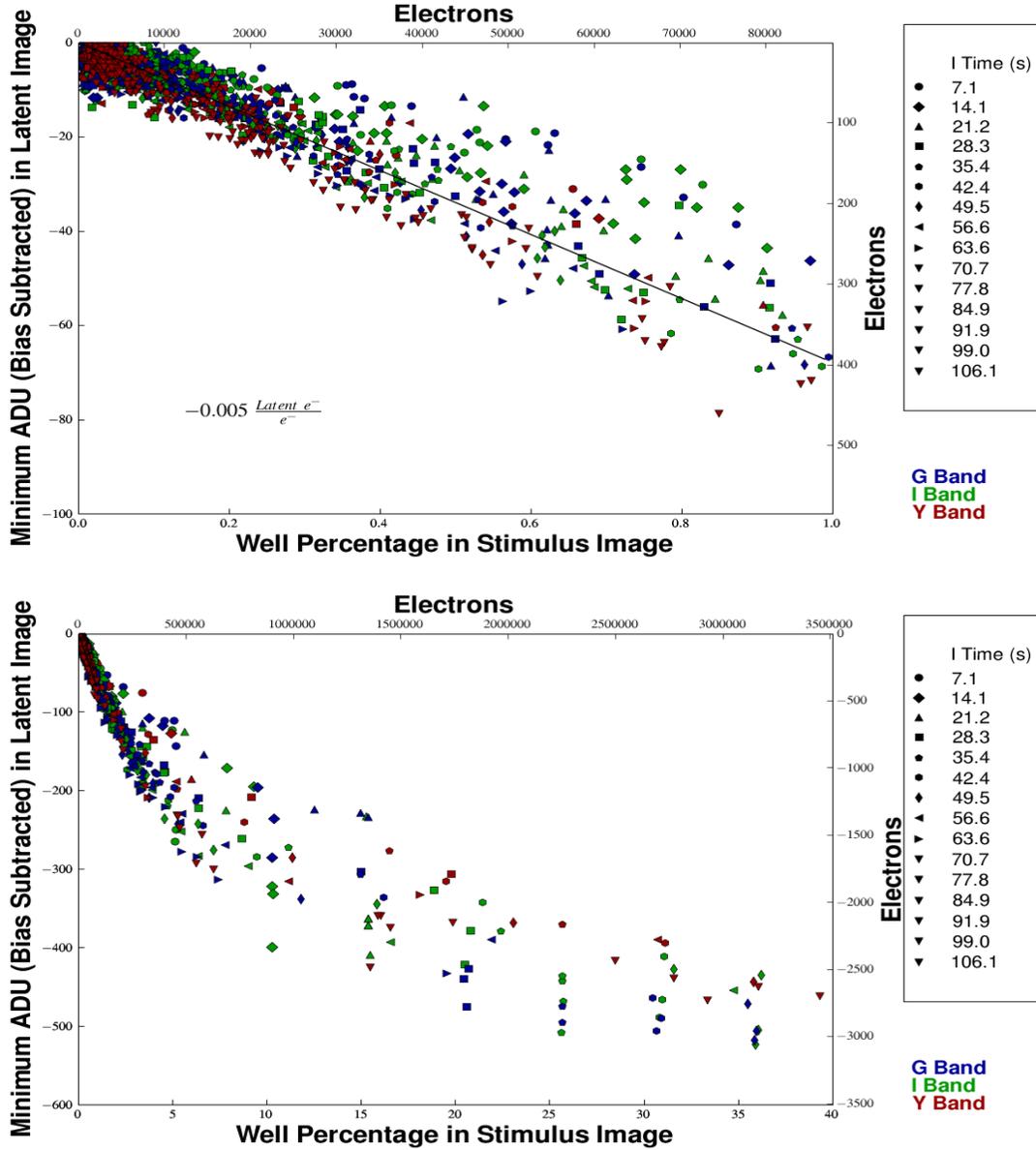


Figure 7.6: These plots show the relationship between persistence signal in the *halo* and the stimulus that caused it. On the *x* axis is the maximum signal attained by the illuminated stimulus (fluence). Values beyond full well are extrapolated by fitting a slope to the pixel before it is saturated. On the *y* axis is the minimum persistence signal in the negative persistence *halo* attained in a 14.4 second dark exposure following the illuminated stimulus. Each data point in these plots corresponds to a particular star in a particular image. The integration time for the point is listed on the legend at the right.

Note that the persistence response for both the core and halo is fairly linear with respect to stimulus from 0 to 100% of the full well (FW). This is proof that the detector pixels need not be near saturation to exhibit latent images. A linear fit yields about 0.027 latent image electrons per stimulus electron at the peak of the core and about 0.005 at the halo minimum. Again, it must be emphasized that the core pixels are seeing an increase in potential and the halo pixels are seeing a decrease in potential. The conversion from potential to electrons is made using the same conversion gain obtained from Fe<sup>55</sup> calibration. After FW, the response becomes nonlinear and at about 10×FW, a saturation of the persistence signal is observed.

Also noteworthy is a slight flux/integration time dependence. In the plot of the maximum signal, for instance, the data points that lie above the line are mostly for exposures longer than 21 seconds and those below the line are for exposures shorter than 21 seconds. For a given well percentage (fluence), the exposure time needed for a star to reach that well percentage is inversely proportional to its magnitude (flux). Therefore, the short integration stars below the line have a higher flux while the longer integration stars above it have a lower flux. The correlation to the filter band is expected since most of the stars in these open clusters have intrinsic blue colors, meaning they are bright in the *g* band and faint in the *y* band.

The larger persistence signal for longer integration times might indicate that the accumulated holes are being trapped at the collecting node while the detector is integrating. To say this with certainty an additional experiment needs to be performed. In this experiment, a pulse of light of duration  $\Delta t$  would be shined on the detector at different times relative to the start and stop of an exposure of duration  $t_{exp}$ , where  $\Delta t \ll t_{exp}$ . If the holes are being trapped after collection, the persistence would be much greater when the pulse occurs near the beginning of the integration than when it occurs at the end.

### 7.2.2.3 Decay Time

The dark current signal due to persistence decays with a timescale that depends strongly on temperature and a number of other factors, which will be considered in the next section. While it can potentially last several hours, for the experiment described in Section 7.2.2.1, the heightened dark current typically decayed to its equilibrium value in 2-4 minutes. The decay appears to be exponential, and is better fit with two time constants (each with a separate coefficient) than one:

$$D_{avg}(t) = D_1 e^{-t/\tau_1} + D_2 e^{-t/\tau_2} + D_{equil}, \quad (7.16)$$

$D_{avg}(t)$  is the instantaneous dark current at time  $t$ , and  $D_1$  and  $D_2$  are components related to the time constants  $\tau_1$  and  $\tau_2$ , respectively.  $D_{equil}$  represents the dark current after the persistence has subsided. Solomon fit persistence in InSb arrays with the same functional form<sup>3</sup> and attributed the

<sup>3</sup>The actual fit was to the number of holes released from traps as a function of time. This can easily be obtained by integrating Equation 7.16 from  $t = 0$  to  $t = \infty$ .

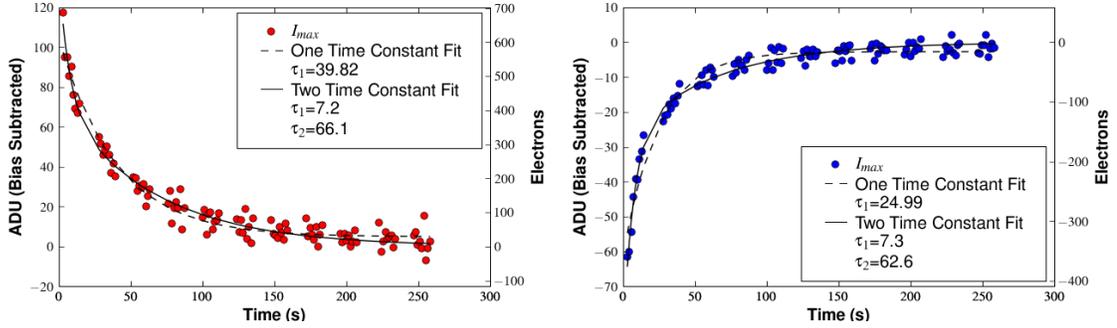


Figure 7.7: (Left) The decay of dark current for  $I_{Max}$  at the core center. In the core, the signal voltage increases with time. (Right) The decay of dark current for  $I_{min}$  at the radius  $R_{halo}$ . In the halo, the signal voltage decreases with time.

two time constants to two energy levels of a single trapping state [42].

Figure 7.7 shows the two time constant fit to the instantaneous dark rate observed after the persistence stimulus. Both the dark signal vs. time at the center of the core-halo structure and that at the radius  $R_{halo}$  are shown. The time constants are very similar for this example, and Figure 7.8 shows that this trend holds for the majority of the latent images that were successfully fitted.

The fact that the time constants for the rise of  $I_{max}$  and fall of  $I_{min}$  are so close in value strongly suggests that charge is being transferred from the core to the halo. If the rising signal in the core and decaying signal in the halo were the result of one species of trap emitting electrons and another species emitting holes, one would expect the emission time constants to be different since they depend on the trap energies and cross sections. It is thus reasoned that the core-halo persistence is a result of charge carriers being exchanged between the pixels in the core and pixels in the halo.

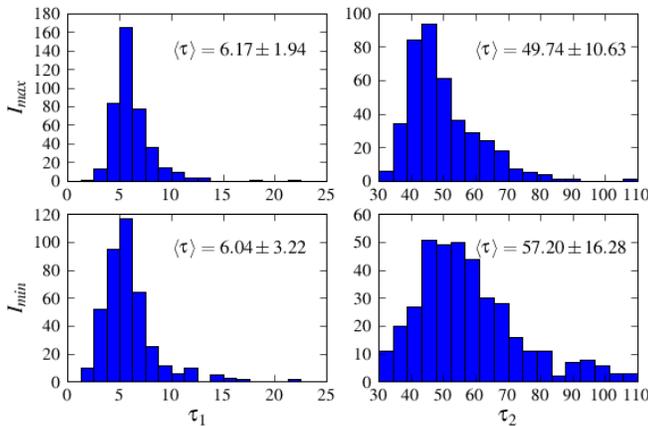


Figure 7.8: Histograms of the time constants for persistence decays of  $I_{max}$  at the center of the latent image and  $I_{min}$  at the radius  $R_{halo}$ .  $\tau_1$  accounts for the fast portion of the decay and  $\tau_2$  accounts for the slower portion. The time constants are in units of seconds.

### 7.2.3 Dependence on Detector Activity

In some cases latent images in the HyViSIs can last far longer than 3-5 minutes. This very long duration persistence is linked to the environmental and operating conditions of the detector, i.e. the temperature, whether the detector is exposed to light or darkness, whether the pixels are left to integrate or are being held under reset, the bias voltages, etc. While no study was performed to explicitly explore the dependence of the latent images on this parameter space, the dominant culprit appears to be a lack of resets on the pixels. The general statement can be made that *more than one reset must be performed to remove latent images in the HyViSIs*. If only one reset of the frame occurs after stimulus and an hour long dark exposure is taken in which latent images are present, the next exposure will show the latent images as well (unless they are extremely weak).

Figure 7.9 shows an example of a latent image that endured for more than an hour. The dark exposures represented in the plot were taken by H2RG-32-147 in window mode with a  $150 \times 150$  window ( $V_{SUB} = 15$  V,  $T=170$  K). Interestingly, at the end of the first dark exposure, shown in the left third of the plot, the dark current in the center of the latent image **decreases to zero**. The

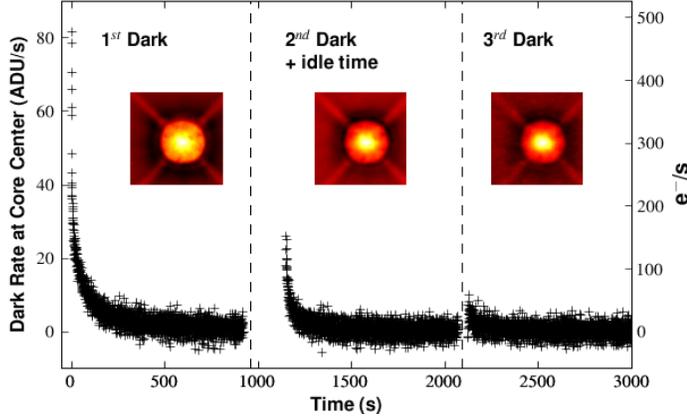


Figure 7.9: An example of a latent image that persisted for more than an hour. The plus marks show the instantaneous dark current measured at the center of the latent image (i.e.  $\Delta I_{max}/\Delta t$ ). The persistence dark current appears to be gone at the end of the first exposure, but reappears in the second.

dark rate in most of the core pixels has done so also. Meanwhile, the pixels at the periphery of the halo continue to show a negative dark current (decrease in voltage) and the outer edge of the halo continues to move radially outward. Surprisingly, after a small amount of idle time following this exposure and one reset, a high dark current in the core pixels reappears in the next exposure. And after an additional reset, it appears again in an exposure taken more than 2000 seconds after the initial stimulus. The latent image thus endures far longer when only two resets are applied than it did in the experiment of Section 7.2.2.1 when resets were being applied every 24 seconds.

**Emission of holes from shallow traps cannot be used to explain this behavior.** If the increase in core signal was a result of holes in shallow traps ( $E_T - E_V \ll 3kT$ ) at the surface or in the bulk being emitted at a rate  $e_p$ , the emission would continue until all of the traps are

empty, regardless of whether or not the pixels are reset or left to integrate (see Section 7.1.1 for an explanation of why this is so). If conventional trap theory is to be used to explain this behavior, it suggests that the traps are deep traps, and that sometime after reset, these deep traps are eventually “starved” of carriers with which to recombine. Additional resets are required to inject the carriers necessary to “feed” these traps and restore the detector to the equilibrium state before the stimulus occurred. A qualitative explanation of how this may occur is pursued in Section 7.5.

### 7.2.4 Dependence on Temperature

The measurements made for this dissertation indicate that persistent charge emission takes place at temperatures from 90 K-200 K. Below this range, it is likely that persistence still occurs; 90 K was simply the lowest temperature that was probed. Above it, the thermal dark current is so high that it completely swamps out the persistent dark current. A thorough study of the relationship between the temporal and spatial characteristics of the latent images and temperature was not carried out due to a lack of time and resources. Nearly all of the latent images in our data were recorded between 160-170 K. However, from the limited sample of data that was collected, several important qualitative aspects stand out.

Figure 7.10 shows the most noteworthy feature of the core-halo persistence at temperatures below 110 K: that is, the halo is almost non-existent. In the image at bottom right, which shows the last minus first read of an exposure taken immediately after an LED light source was removed from the H2RG-001 at 100 K, the halo is nearly undetectable because its amplitude is much smaller than the range of pixel values in the stretch. In the plot at left, which shows a plot of the pixel values along the column highlighted by the cyan line in the image, one can see that the halo is a small dip that occupies 1-2 pixels at the edge of the latent image. The pixels in the halo lose only 1-2 mV of signal in 14 seconds at this temperature. Meanwhile, the pixels in the core gain about 35 mV of signal, which is comparable to the amount of signal gained in 14 seconds by the core pixels in a 160 K latent image. Thus, the behavior of the halo pixels is extremely different in these two temperature extremes, while that for the core pixels is somewhat similar.

Although the core pixels share a common persistent dark rate at these temperature extremes, Figure 7.10 reveals a difference in the way they act. The core pixels near the center of the latent image (which has a radius of about 600 pixels at 100 K and a radius of about 800 pixels at 180 K) form a rather flat pedestal in signal at any time,  $t$ , for both temperatures. In other words, they are all seeing the same dark rate,  $D(t)$ , over the exposure time. At 100 K there is a very small difference in the dark rate at any location in the core, and it looks essentially flat. But at 175 K, near the core-halo boundary, the pixels in the halo see a much greater  $D(t)$  than the ones near the center, resulting in the peak in the left plot, and the pixels immediately outside see a negative  $D(t)$ , resulting in the trough. One cannot help but reason that the formation of the peak and trough is due to the horizontal exchange of charge carriers somewhere in the detector.

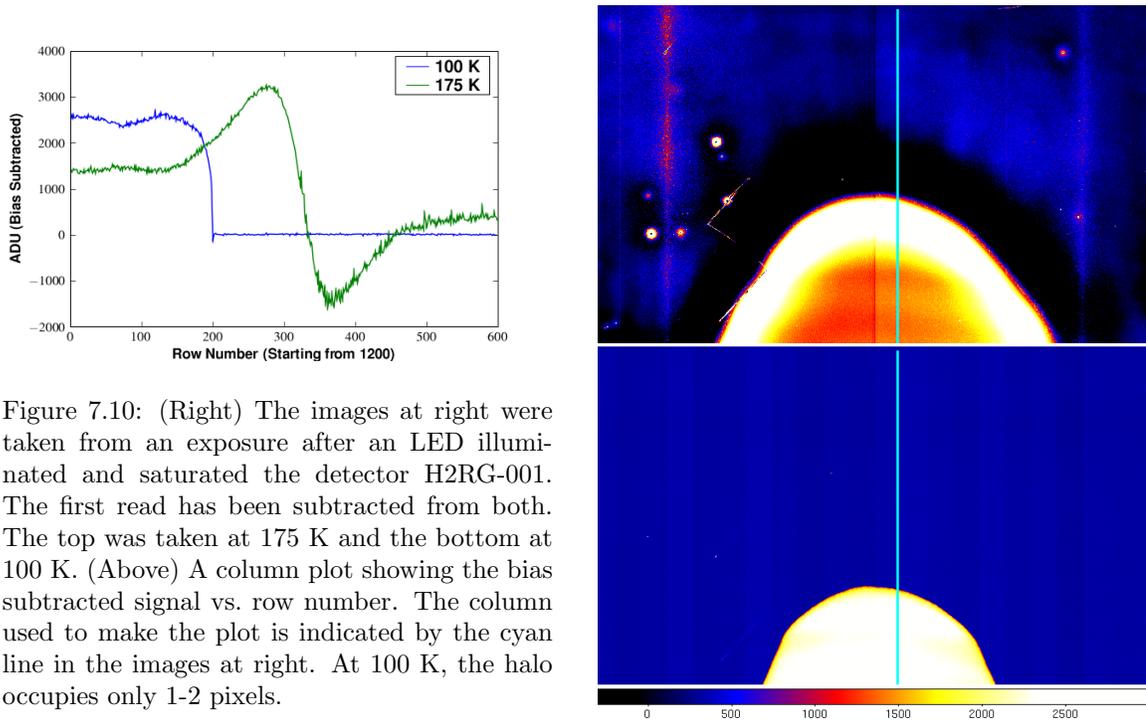


Figure 7.10: (Right) The images at right were taken from an exposure after an LED illuminated and saturated the detector H2RG-001. The first read has been subtracted from both. The top was taken at 175 K and the bottom at 100 K. (Above) A column plot showing the bias subtracted signal vs. row number. The column used to make the plot is indicated by the cyan line in the images at right. At 100 K, the halo occupies only 1-2 pixels.

Nearly all properties of silicon—the carrier mobilities and diffusion coefficients, intrinsic carrier concentration, resistivity, etc.—are dependent on temperature. But if we believe that charge transport through the detector material is responsible for the halo and peak near the core edge, then the most likely candidate variable is the diffusion coefficient for electrons,  $D_n$ . Electrons injected into a pixel at reset do not see a large inter-pixel potential barrier if the front surface between  $p^+$  implants is in an accumulated state, so they are free to diffuse. For pixels near the center of the large core, there is no gradient in the concentration of these electrons, and thus no diffusion to cause a discrepancy in the dark rates. A change in signal level for these pixels still occurs as electrons diffuse toward the back surface, though. Near the edge of the core, where the pixels outside have not been filled with electrons, there is such a gradient, and the diffusion adds an additional component to the signal change. At low temperatures, the diffusion of these electrons may occur too slowly to cause an appreciable decrease in the halo or let it reach an appreciable distance. At higher temperatures (the diffusion length should go as  $T^{3/4}$  [117]), the opposite is true. The core-edge pixels see a larger signal rise because of the electrons diffusing horizontally out of them and the halo pixels that receive these electrons see a decrease. This theory will be explored further in Section 7.5.

### 7.2.5 Dependence on $V_{SUB}$

The latent images show very interesting spatial signatures as  $V_{SUB}$  is varied, especially below 8 V, when the detector is not fully depleted. As can be seen in Figure 7.11, the spatial distribution of the photoholes collected during the stimulus image are equally interesting. This figure encapsulates a qualitative summary of a very long imaging process performed at the telescope with H1RG-022 and warrants a thorough description.

#### 7.2.5.1 Description of Latent Image Exposure Sequence

Each of the image snippets in Figure 7.11 shows the fourth exposure out of a four exposure sequence at a particular value of  $V_{SUB}$ , listed in green. They were all recorded with H1RG-022. In this fourth exposure, the star being imaged is in the upper left quadrant. In the third exposure the star was imaged in the upper right quadrant, in the second it was imaged in the lower right quadrant, and in the first it was imaged in the lower left quadrant. For values of  $V_{SUB} > 1.5$  V, latent images are evident in all three of these quadrants. The latent image in the upper right quadrant has the “freshest” persistence since it was stimulated 71 seconds prior to the fourth exposure and the one in the lower left is the “least fresh” since it was stimulated 213 seconds prior. The image snippets were made by subtracting the last read of a 100 read exposure from the first (bias subtraction). No dark current subtraction or flat field was applied. The telescope was dithered by about two arcminutes between exposures to bring the star to its new location. Idle resets were performed at  $V_{RESET} = 0.094$  V in between exposures and the temperature was 170 K.<sup>4</sup> The detector was operated in 16 output mode with a frame time of  $t_{frame} = 0.71$  s.

#### 7.2.5.2 Charge Collection

It is immediately clear from Figure 7.11 that charge collection in the pixels depends strongly on the backside voltage. The diffusive behavior in the 1.5-3.0 volt range agrees well with the model presented in Section 2.2.2.2 in that it shows fairly radially symmetric diffusion when the depletion regions collapse. For  $V_{SUB} < 1.5$ V and  $V_{SUB} > 3.0$ V, on the other hand, there are substantial deviations from radial symmetry that are difficult to overlook.

**For  $V_{SUB} \leq 1.5$  V** a clear pattern following the columns in the multiplexer is evident. The pixels along the odd multiples of 64 (OMC) see a rapid decrease in voltage of  $\sim 50 - 60$  mV in about 10 seconds. After the drop, they either begin to integrate or hold steady. The pixels along the even multiples (EMC) see a quasi-exponential increase of  $\sim 120 - 190$  mV over the full 71 seconds. For pixels between two OMCs, the change in voltage over the 71 seconds increases monotonically from

<sup>4</sup>If resets were not performed while the telescope was moved, the star would leave a latent image *trail* in its path along the detector. These trails are present throughout the telescope data from the instances when no resets were performed while the telescope was being slewed to a new field.

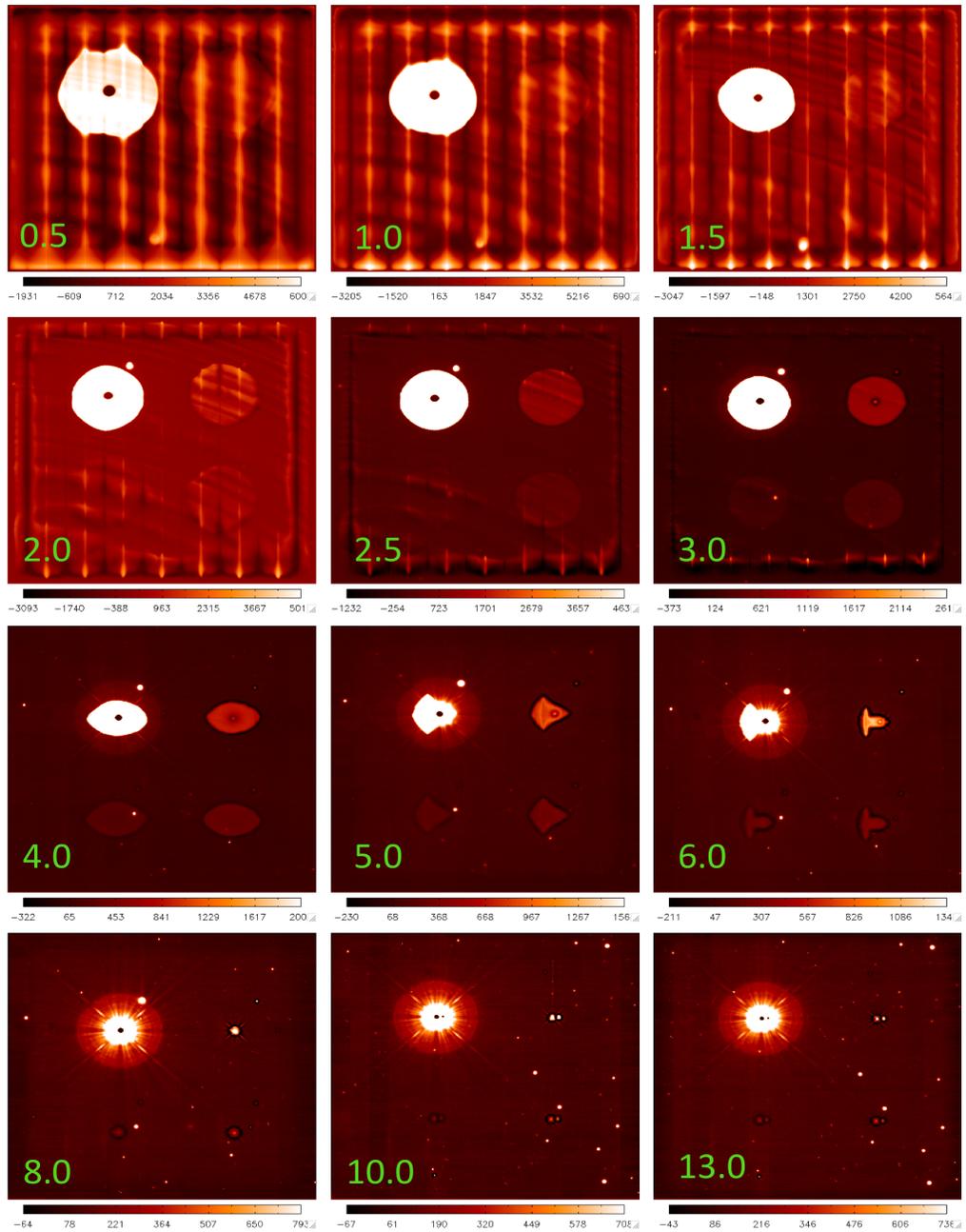


Figure 7.11: Latent images for saturating stars at different values of  $V_{SUB}$  taken through  $i$  band. The value of  $V_{SUB}$  is listed in each image snippet in green. See the text for a description of the imaging sequence.

–60mV to 190 mV and then decreases back down to –60mV along the row. This behavior suggests a horizontal gradient in the detector front surface potential, which might arise because of a voltage differential in the multiplexer. The physical layout of the multiplexer is unknown, so determining the specific cause is difficult. Some possibilities are coupling to current carrying lines that are located at different depths beneath the plane of the detector or to the currents going through the 16 outputs (the schematic in the manual indicates that the individual column buses carry the signals off chip before they are actually routed to the outputs, though, so the latter may not be realistic; also, the pattern is present even when the detector is run through one output).

Charge carriers generated by the star diffuse radially outward, but advance more quickly along the EMC and less quickly along the OMC. While this behavior is interesting, understanding it is of little value for astronomical application. Operating the detector at such low bias voltages would yield extremely poor performance.

**For  $1.5 \text{ V} < V_{SUB} \leq 3.0 \text{ V}$**  the observations match very well with the expectations outlined in Section 2.2.2.2. After the depletion regions of the illuminated pixels collapse, the carriers diffuse radially outward and destroy the depletion regions in a new annulus of pixels, creating a domino effect of blooming.

**For  $3.0 < V_{SUB} < 8.0 \text{ V}$**  a clear asymmetry in the diffusion is evident. At 4.0 V, the carriers diffuse more strongly in the x direction, resulting in a highly elliptical star. At 5.0 and 6.0 V, they diffuse more strongly along the x direction as well. Oddly, though, another asymmetry is observed between the  $-/+x$  directions at these voltages. Once carriers diffusing in the  $-x$  direction arrive at one of the OMCs, they show a strong tendency to follow the OMC vertically. Note that this is opposite to the behavior for  $V_{SUB} < 1.5 \text{ V}$ , where the charge followed the EMCs. In the  $+x$  direction, the charge cloud tightens as it moves. The most logical explanation for this is that the front-side potential is lower along the OMCs than it is along the EMCs at these bias voltages. The holes thus tend to sink into these wells more easily.

It is not at all clear why the disparity in surface potential is so strong when  $V_{SUB}$  is between 5-6 V or why such a large asymmetry sets in at 4 V. Understanding this would require a two or three dimensional simulation on top of intimate knowledge of the detector and multiplexer layout. A simulation of this nature would be useful, though, since *these asymmetries might very well offer a clue into the  $x - y$  asymmetry exhibited in pixel crosstalk and interpixel capacitance* (see Sections 6.1.4 and 6.1.3). With or without comprehension of this phenomenon, the “point spread functions” at these voltages demonstrate that the detector should be operated at  $V_{SUB} \geq 8 \text{ V}$  for astronomical applications.

**For  $V_{SUB} > 8.0 \text{ V}$**  the detector achieves more ideal performance as it enters the overdepleted state. There is no sign of the depletion regions collapsing and allowing photocharges to invade

neighboring pixels. Bright stars still generate significant photocurrent outside the FWHM, but this is a result of diffracted light and not diffusion of carriers from field free regions.

### 7.2.5.3 Charge Persistence

Perhaps the most striking attribute of latent images generated when the detector is not fully depleted is *the presence of two core-halo structures (CHS)*. This is most easily seen in the  $V_{SUB} = 3.0$  V image of Figure 7.11. One CHS, call it  $CHS_{inner}$ , emerges near the center of the image, with  $r_{core} \sim FWHM$  of the star in the previous exposure and  $r_{halo} \sim 2 - 3 \times FWHM$ . Another, call it  $CHS_{outer}$ , emerges at the outskirts of the charge distribution created by the star in the previous exposure. In between these two regions, the pixels show roughly the same increase in signal due to persistence, making it one large core, effectively.

Figure 7.12 provides a key into understanding how this happens. The observations to this point imply that persistent signal arises because of carriers trapped near the  $p^+-n$  junction, i.e. near the front surface (the way in which these trapped carriers actually *generate* persistence will be discussed in Section 7.5). If the number of trapped carriers is proportional to both the number of stored majority holes in the collecting node **and** the excess minority holes in the bulk, then the trapped carrier distribution should look similar to the one shown in Figure 7.12. One can see that the locations of the  $CHS_{inner}$  and  $CHS_{outer}$  are seen to coincide with the steep gradients in hole concentrations at  $r_1$  and  $r_2$ , respectively. It was already shown in the previous section that diffusion is a likely candidate for the charge transport that generates the CHS. The existence of CHSs in regions where there is a strong gradient  $dp/dr$  in carrier concentration provides further support for this. And when the field free diffusion regions vanish in overdepletion,  $r_2$  is drawn in and merges with  $r_1$  since  $\Delta p_{diff} = 0$ . This explains why  $CHS_{outer}$  vanishes as  $V_{SUB}$  is increased.

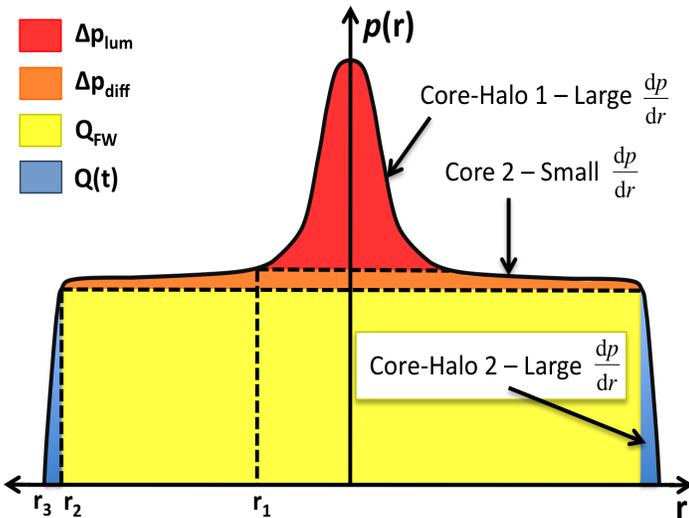


Figure 7.12: Number of stored and free holes versus radius when light from a saturating star is incident on the undepleted HyViSI. Following Section 2.2.2.2, the stored holes,  $Q_{FW}$  and  $Q(t)$  have accumulated in the  $p^+$  implant while the free holes,  $\Delta p_{lum}$  and  $\Delta p_{diff}$  exist as excess minority carriers in the field free regions near the metallurgical  $p^+-n$  junction. At reset, the large gradients in these holes translates to large gradients in injected carriers, and core-halo persistence occurs at these gradients.

### 7.3 Semi-Permanent Offsets: Laser Burn-In

There is another type of “after-image” that can be generated in HyViSI devices, which is of a fundamentally different nature than the short-term persistence just described. We refer to this effect as **Laser Burn-In** (LBI) because it has been accidentally discovered by several groups who used a laser to illuminate the detector in laboratory experiments. Of course, any light source that produces a beam of comparable intensity to a laser should create such an after-image. Figure 7.13, which shows two offsets created by imaging Mars through the 2.1m telescope with H2RG-32-147, proves this is indeed the case.

The unique thing to note about Figure 7.13 is that it is a reset frame of the detector, meaning that the pixel values are read while the reset switches are held closed in the pixels of the multiplexer. This is very convincing evidence that the origin of the offset lies in the ROIC and not the detector. Bai et al. [25] provide confirmation of this, attributing the offset to “damage” to the gate oxide of the pixel source follower, which has a minimum rated breakdown voltage of  $\sim 5.2$  V (damage is placed in quotes because the offset will eventually anneal after a very long time, as will be shown). In infrared hybrid CMOS detectors, the effect is not observed because the small reverse bias used ( $< 1$  V) does not allow the sense node to get this high. But with the large backside voltage needed for astronomy (15-50 V), the pixel source followers in HyViSI sensors are susceptible to this damage.

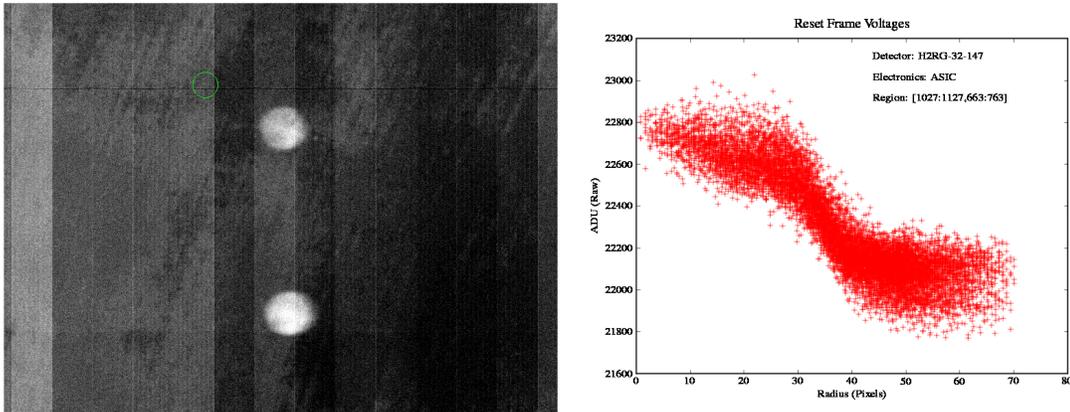


Figure 7.13: (Left) Image of raw ADU values in a reset frame showing the offset induced by imaging Mars through the Kitt Peak 2.1m telescope. (Right) The radial profile shows the elevation in ADU values relative to the background. The peak to valley of  $\sim 700$  ADU corresponds to about 45 mV.

The conditions under which the LBI occurs, referred to by Bai et al. [25] as *supersaturation*, are a complicated function of actual breakdown voltage, detector bias, temperature, input light level, integration time, and the time between resets of the sense node. The empirical evidence indicates that the biggest of these factors is the amount of flux falling on the detector. For instance, a short, 10 second exposure with a 5 pixel FWHM laser beam at a high flux of  $3.54$  kW/m<sup>2</sup> and  $V_{SUB} = 7$  V

was enough to create a 50 mV offset at room temperature in a HyViSI H2RG. But for long duration exposures of bright star fields, where the fluence of a given star was  $50\text{-}100 \times$  full well with a low flux of  $< 10 \text{ W/m}^2$ , no offset was created at  $V_{SUB} = 15 \text{ V}$ . The reason for this is that there are leakage paths (such as the thermionic diffusion discussed in Section 6.1.4.3) to remove the accumulating holes and prevent the voltage at the sense node from getting as high as 5.2 V in the low flux case. It is only when the amount of photocurrent generated above the pixel is sufficiently large that it cannot be effectively removed through the leakage paths where the voltage can rise above 5.2 V. This happens with high flux conditions.

The exact relationship between the operating conditions and the resulting offset is not known, nor is the maximum voltage offset that can be created.<sup>5</sup> In the frame shown in Figure 7.13, the offset is approximately 45 mV. In other laboratory LBI, offsets as large as 120 mV have been induced. It is possible the upper limit for the offset may extend even further. Fortunately, **the LBI offset is purely DC and has no measurable effect on photocurrent estimates using CDS, Fowler Sampling, or slopefitting**, which are the methods used to measure signal in astronomical applications. Unfortunately, the offset reduces the practical well depth in the affected pixels, and hence, their dynamic range. For instance, immediately after observing Mars with H2RG-32-147, the stressed pixels started at a level 45 mV (800 ADU at a gain of 1 on the SIDECAR) above the level measured prior to burn-in, but still saturated at the same level as the one measured prior to burn-in. Thus, their effective well depth is reduced by about  $4500 e^-$ . Several tests were run with adjusted ADC ranges for the readout electronics to confirm that the detector output is saturated and not the ADC.

It was originally speculated that the LBI was a form of permanent damage to the sensor. As Figure 7.14 shows, though, the offset does anneal over time. This can take anywhere from days to several months. For the particular case shown in Figure 7.14, the spot decayed over six months while the detector was kept at room temperature and used somewhat infrequently. The Mars burn-in showed no signs of a reduction over three months time with the detector cooled between 100-200 K and used regularly. While three to six months may seem like an extremely long time, previous studies have shown that in Si-SiO<sub>2</sub> interfaces, traps in SiO<sub>2</sub> can exchange charge with a Si substrate on time scales ranging from less than 1  $\mu\text{s}$  to many years [118]. In fact, many electrically programmable memory devices rely on such a longevity of trapped charge in oxides to store data [119]. There is some indication that constantly resetting the detector can aid in disintegrating the spot over a much quicker time, though. A 120 mV offset burned into H2RG-148 with a  $1\text{ kW/m}^2$  laser disappeared over 10 days. The detector was held at 160K and  $V_{SUB} = 15 \text{ V}$  for the majority of this time with idle resets being constantly applied to the detector while it was not taking exposures.

Regardless of the time required for the LBI to vanish, the loss in well depth is obviously a concern for astronomical imaging since the largest possible dynamic range is desired. To avoid

---

<sup>5</sup>There is some degree of apprehension involved with the LBI since it could potentially compromise these very expensive devices.

damage, Teledyne Imaging Sensors recommends either 1) biasing the detector with  $V_{SUB} < 5$  V or 2) avoiding oversaturation. The first of these is highly impractical for astronomical applications, since the PSF at such a low bias voltage is quite large and takes on the strange shapes shown in the previous section. Also, the dark current volcanoes occupy a much larger radius, rendering many pixels unusable when the detector is not fully depleted. The second of these might be a potential solution, although it will truly be a challenge to implement. It requires knowing in advance the position of bright stars and planets and using the guide mode feature of the detector to reset the pixels subject to the most concentrated light. Resetting the pixels will prevent supersaturation and the ensuing damage, with the side effect of creating offsets along the columns and rows of the guide window (see Section 6.1.4.3).

In certain optical systems, the size of the telescope aperture, filters used, plate scale, etc. might preclude the LBI from presenting a threat at all. For instance, in our three weeks of 2.1m observations, in which we observed many stars between 1<sup>st</sup> and 5<sup>th</sup> magnitude, the only source that caused the offset was Mars. In fact, imaging the planet through  $g$ ,  $i$ , and  $y$  did not create an offset. It was only when no filter was used that the damage was done. An accurate calculation of the flux that Mars produced at the detector when no filter was present based upon its apparent magnitude is not possible. The detector was saturated before the first read, so estimating the flux from photocurrent is not an option, and the transmission of the mirrors and detector window are not known, so using the magnitude to obtain a flux at the detector would be equally uncertain. It is likely that it was somewhere on the order of  $1\text{ kW/m}^2$  like the lasers that produced LBI in the laboratory.

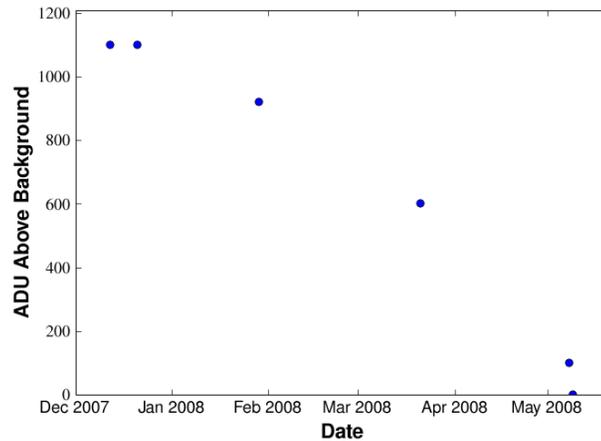


Figure 7.14: A plot showing the average LBI offset in H2RG-32-029 over a 6 month period. The spot was burned in by a monochromatic laser with a 650 nm wavelength. At best focus the beam produced about  $3.54\text{ kW/m}^2$ . The detector was kept at room temperature throughout the six months. The plot shows that the spot eventually disappeared. The gain was approximately  $6\text{ e}^-/\text{ADU}$ .

## 7.4 Effect of Forward Biasing

In his studies on persistence in infrared hybrid CMOS detectors with an InSb detector layer, Solomon noted that forward biasing the photodiodes effectively removed latent images [42]. He found that while the latent image could no longer be detected when the reverse bias was restored, the entire detector suffered from an instability in dark current. Similarly, Smith et al. demonstrated that the rise in dark current after incrementing the reverse bias across photodiodes in HgCdTe arrays is almost identical to the rise in dark current after exposure to a light source [116].

In HyViSI detectors we find a nearly analogous behavior. Forward biasing the photodiodes by bringing  $V_{RESET}$  higher than  $V_{SUB}$  and resetting the pixels clears any spatial signature of the latent image. In normal operation, this amounts to decreasing the output voltage on the power source that supplies  $V_{SUB}$  since  $V_{RESET}$  should not be brought above 1.7 volts. Unfortunately, when we reverse bias the diodes again and take another exposure we find the dark current has risen by several orders of magnitude and takes on the same exponential behavior as the persistence signal.<sup>6</sup> Effectively, **every pixel is afflicted with persistence** after the forward bias. It requires minutes to hours for the dark current to settle nonlinearly back to its equilibrium value.

In the following sections, we present an overview of the theory behind switching from forward to reverse bias in a PIN diode. We then describe experiments which show that the dark current signal behavior—the characteristic timescales and magnitude—are essentially the same in latent images and after forward biasing. Lastly, we show that varying the number of resets and the readout sequence after the reverse bias has been restored does not alter this behavior.

### 7.4.1 Theory of Bias-Direction Switch

The time required for the dark current to settle to equilibrium after going from forward to reverse bias is a major consideration in the design of PIN diodes for switching applications [54]. When a PIN diode is forward biased, majority carrier electrons are injected into the intrinsic or  $\nu$ -type region from the  $n^+$  region and majority carrier holes are injected into the  $\nu$ -type region from the  $p^+$  region (see Figure 2.1). The greater rate of recombination of carriers when holes from the  $p^+$  side meet the electrons from the  $n^+$  side over generation leads to an accumulation of charge in the  $\nu$  region. Excess carriers also accumulate in the heavily doped regions. To reach equilibrium in the reverse bias state, these injected carriers must be swept out of the diode to form the depletion region.

Simulations and theoretical calculations show that the heavily doped contacts are depleted quickly (on the order of  $1/20^{th}$  the lifetime of the minority carriers) relative to the intrinsic, or lightly doped  $\nu$  region [120, 121]. The time it takes for the latter to occur is directly related to the width of the diode  $W$ , and is unaffected by the minority carrier lifetimes as long as the exterior circuitry permits large sufficiently large reverse currents [121]. Interestingly, in PIN diodes used for

---

<sup>6</sup>Unless a distinction is necessary, the process of forward biasing the diodes and returning them to their original reverse bias value will be referred to as a *bias switch* for the sake of brevity.

power switching applications, this time is on the order of milliseconds or microseconds. For the HyViSI PIN diodes, it is on the order of minutes, suggesting that another mechanism is responsible for the long time taken (or that the reverse current cannot reach some spots in the detector layer).

## 7.4.2 Forward Bias to Full Reverse Bias of $V_{SUB} = 15\text{V}$

After observing the removal of latent images with a forward bias, one is inclined to ask: *does this method offer a better solution for persistence mitigation than simply waiting for the latent images to subside?* Further, *does varying the time the pixels are held in reset versus the time they are allowed to integrate affect the time required for the detector to return to its equilibrium state?* An experiment was carried out to answer these questions and gain a better understanding of the mechanisms at play with latent images.

### 7.4.2.1 Experimental Description

The experiment is carried out as follows:

- 1) After the detector dark current has reached its equilibrium value, an exposure is taken consisting of 1 reset (line by line) and 10 reads at  $V_{SUB} = 15\text{ V}$  and  $V_{RESET} = 300\text{ mV}$ . Line by line reset is used.
- 2) The backside voltage is brought down to  $V_{SUB} = 0\text{ V}$  and another exposure with 10 (line by line) resets is taken at  $V_{RESET} = 300\text{ mV}$ . The PIN diodes are forward biased while the reset is held down and the pixel values are read out while the resets are performed.
- 3) The backside voltage is brought back to  $V_{SUB} = 15\text{V}$ . Exposures are then taken with a cadence of  $N_{Resets}$  and  $N_{Reads}$  while maintaining  $V_{RESET} = 300\text{ mV}$  until the dark current returns to thermal equilibrium.
- 4) Steps 1-3 are repeated with a different cadence.

All data was taken with the H1RG-022 at 170 K in 16 output full frame mode. The power supply that sourced  $V_{SUB}$  was adjusted through an RS-232 interface by the same IDL script that controlled the imaging sequence, which provided precise, uniform timing for each operation. Approximately 34 seconds elapsed between both the drop in  $V_{SUB}$  from 15 to 0 V and the forward bias reset to the return of  $V_{SUB} = 15\text{V}$ . The voltage was stepped in increments of  $\Delta V_{SUB} = 5\text{ V}$ . Based on a similar sequence used with H2RG-32-147 to erase latent images in telescope observations, the results appear to be applicable to all HyViSI detectors.

### 7.4.2.2 Results

Figure 7.15 shows the average dark current measured after step **3** of the experiment for several different values of  $N_{Resets}$  and  $N_{Reads}$ . The dark current at each time (read) was measured by

differencing the pixel values in successive reads of the detector, averaging a  $150 \times 150$  box of pixels, and dividing by the time between reads. The main result highlighted by the figure is: **the return of the dark current to equilibrium after forward biasing the diodes and restoring reverse bias is not influenced by the number of resets or number of reads performed.** In addition, the sequence of  $N_{Resets} = 1$  and  $N_{Reads} = 250$  was performed with and without drop frames, indicating that leaving the detector idle yields the same result.

The average dark current as a function of time,  $D_{avg}(t)$  (in  $e^-/s/pix$ ), is well fit by an exponential decay of the form

$$D_{avg}(t) = D_1 e^{-t/\tau_1} + D_2 e^{-t/\tau_2} + D_{equil}, \quad (7.17)$$

where  $\tau_1$  and  $\tau_2$  are time constants associated with dark current amplitudes  $D_1$  and  $D_2$ , respectively, and  $D_{equil}$  is the average dark current after the return to equilibrium. The average value of the coefficients from fits to the five readout sequences are:

$$\begin{aligned} \tau_1 &= 13.1 \text{ s} & \tau_2 &= 91.1 \text{ s} \\ D_1 &= 244.4 \text{ e}^-/s/pix & D_2 &= 121.17 \text{ e}^-/s/pix \\ D_{equil} &= 11.5 \text{ e}^-/s/pix \end{aligned}$$

The fit with these average values is shown in the black curves of Figure 7.15. While the fit describes the data well for  $t = 0$  to  $t = 400$  s, a very slow decay continues to occur for  $t > 400$  s and brings  $D_{equil}$  to about  $7.5 \text{ e}^-/s/pix$ , suggesting a third time constant may be involved. A similar long time constant was observed for the latent images in InSb arrays [42].

The time constants associated with the decay are slightly longer than the time constants associated with the persistence signal. Possible reasons for this will be discussed in Section 7.5. If the source of the two phenomena are the same, then the total amount of carriers released by a given pixel en route to equilibrium after the bias switch should still be the same as amount of carriers released by that pixel after it has reached saturation in persistence. The total amount of carriers can be calculated by integrating Equation 7.17 from  $t = 0$  to  $t = \infty$ :

$$N_1 = D_1 \tau_1 \cong 3200 \text{ e}^- \quad N_2 = D_2 \tau_2 \cong 11,000 \text{ e}^- \quad (7.18)$$

The sum of these two terms is about  $14,000 \text{ e}^-$ , which is very close to the maximum signal rise observed in a latent image with H1RG-022. The physical explanation of these these time constants and coefficients, as well as the relationship of the dark currents after bias switching and latent images will be discussed further. First, a discussion of effects seen after incrementing  $V_{SUB}$  in steps is given to shed light on where in the detector the dark carriers are originating.

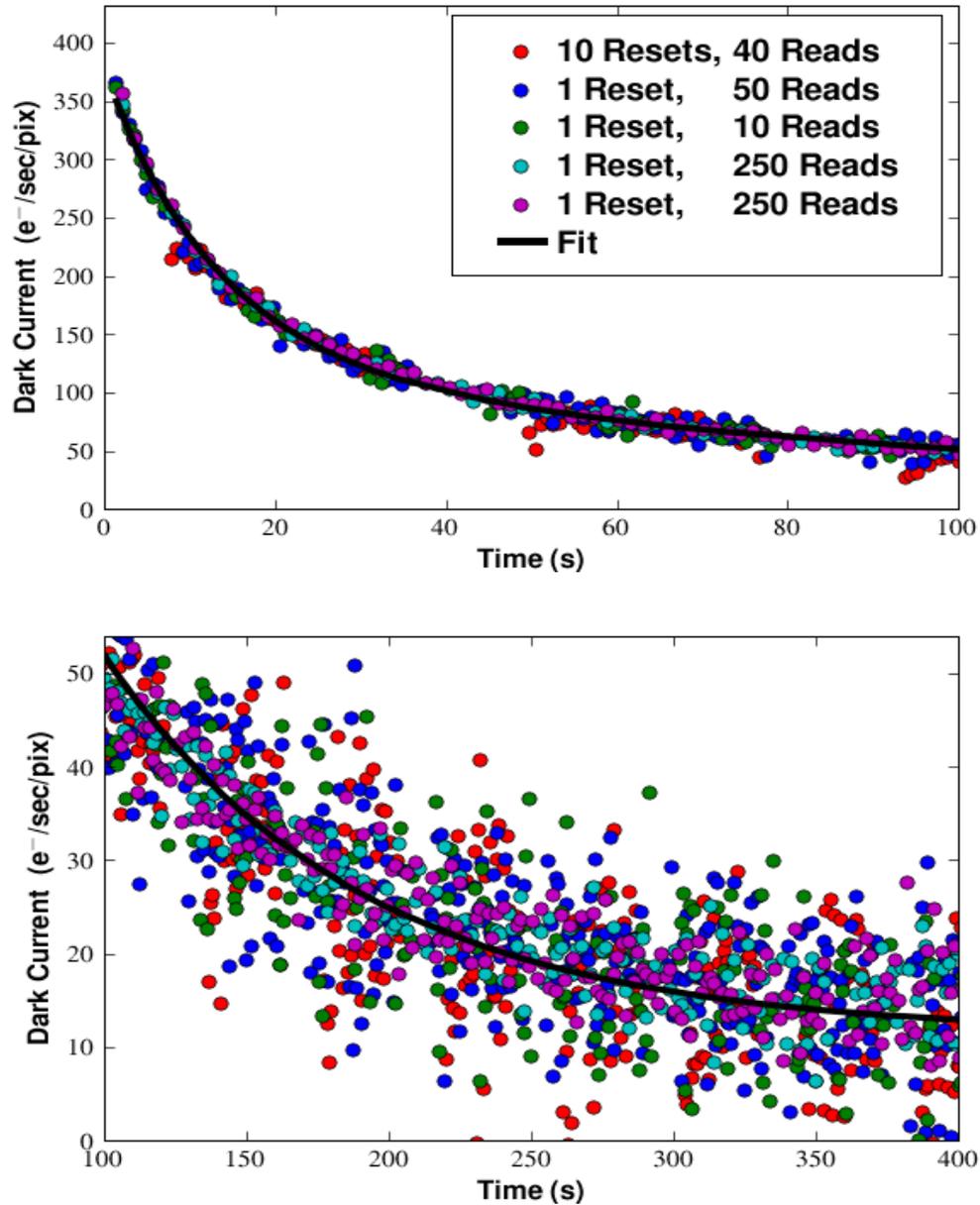


Figure 7.15: Dark current after forward biasing the PIN diodes and then restoring the reverse bias. The number of reads in an exposure and the number of resets between them is indicated in the legend. The outlying points for  $N_{Resets} = 10$  and  $N_{Reads} = 40$  is caused by a droop in signal that was not fully corrected by reference pixel subtraction. The fit from Equation 7.17 is shown with the black curve.

### 7.4.3 Incrementing Reverse Bias with $\Delta V_{SUB} = 2\text{V}$

A similar type of procedure to the one described in the previous section was carried out at the Independent Detector Testing Laboratory by Don Figer and colleagues with H1RG-018. Instead of switching from forward to reverse bias, the backside voltage was incremented in steps of  $\Delta V_{SUB} = 2\text{V}$ , starting at  $V_{SUB} = 0\text{V}$  and going up to  $10\text{V}$ . At each voltage, 20 or more dark exposures with 33 reads and an exposure time of 200 seconds were recorded. The detector was held at a temperature of  $160\text{K}$  and readout in full frame mode. No record of the number of resets between exposures or exact timing between them was available from the FITS headers for this data, so it will only be used to provide a qualitative look at the dark current after incrementing  $V_{SUB}$ .

Figure 7.16 shows the average dark current (measured by fitting slopes to the ramps of pixels and taking an average of a  $150 \times 150$  box) as a function of time after each increase in  $V_{SUB}$ . It is immediately clear that the time constant associated with the decay to equilibrium for the steps of 0 to 2 V and 2 to 4 V is much greater than it was for the case when the bias was ramped from 0 to 15 V. Further, it is observed that nearly all the decay occurs between 0 and 4 V, with only a very small change in dark current for the increments in the 4 to 10 V range.

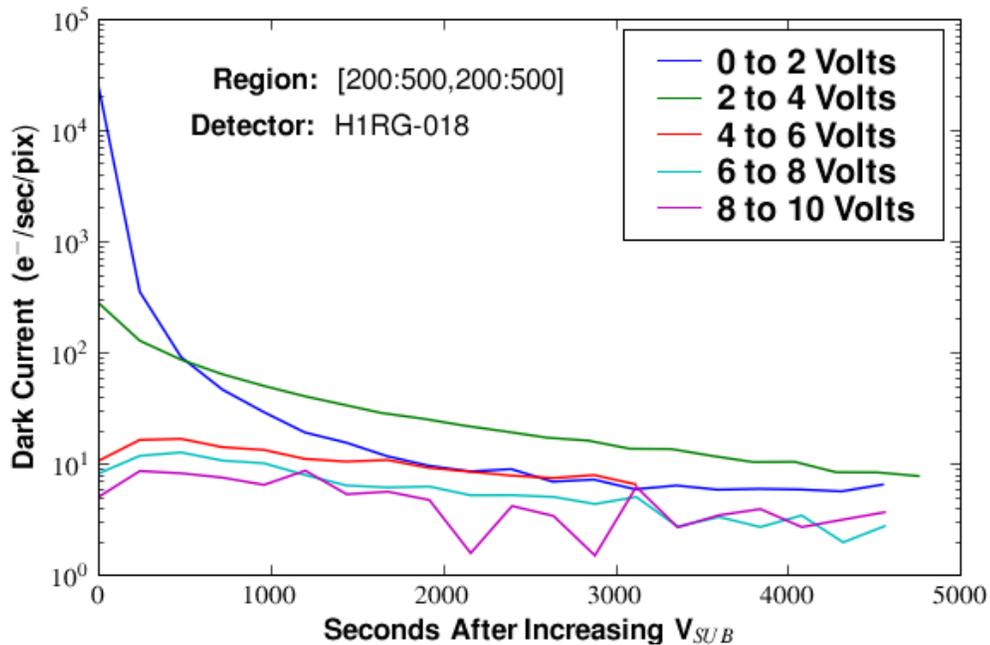


Figure 7.16: The mean dark current in 33 read/200 second integrations after increasing  $V_{SUB}$  in 2 volt increments. The dark current is estimated in each integration by fitting a slope to the pixels in a subregion of the array and taking the mean.

The dependence of the dark current on the change in  $V_{SUB}$  is evidence that the dark carrier generation sites are located near the front surface. If they were located uniformly throughout the bulk along the  $z$ -axis (from the frontside to backside), then when the diodes are not fully depleted, increasing the reverse bias should sweep out a new set of carriers from the layer of material that was previously field free, which would result in a large dark signal. The results of the numerical simulations shown in Figure 2.6 indicate the change of the width of this field free layer goes from 24  $\mu\text{m}$  for  $0 \rightarrow 2$  volts to 20  $\mu\text{m}$  for  $2 \rightarrow 4$  volts and continues to get smaller as  $V_{SUB}$  is incremented higher, reaching about 10  $\mu\text{m}$  for  $8 \rightarrow 10$  volts. If the dark carrier generation sites were indeed located uniformly throughout the diode, one would expect to see a difference in the dark rates when the bias voltage is switched from  $4 \rightarrow 6$ ,  $6 \rightarrow 8$ , and  $8 \rightarrow 10$  volts. One can see from Figure 7.16 that this is not so. Thus, we conclude that the dominant dark generation sites after the switch from forward to reverse bias, which are also believed to be the hole storage sites that participate in persistence, are concentrated within 20-40  $\mu\text{m}$  of the front surface.

Another important consideration is that the metallurgical junction is a two dimensional surface. When  $V_{SUB}$  is raised from  $0 \rightarrow 2$  V, the volume of bulk material in between the  $p^+$  implants, above the front surface is also being swept of carriers. Stored holes that cross the junction into the  $p^+$  implant will cause a large dark current. The one-dimensional numerical simulation does not treat this region, and the two dimensional geometry of the junction is not known. It is very likely that a large contribution to the dark rate arises in this region due to trapped minority carriers being swept out when the reverse bias is applied. When  $V_{SUB}$  is increased further, this region is already near or at full depletion, depending on the potential above the surface passivation. If it is fully depleted, it will not contribute to the dark rate. This may explain why the change in dark rate is so large when  $V_{SUB}$  is increased in the 0-4 V range, and near zero when increased in the 4-10 V range.

## 7.5 Model, Simulations, and Summary

Even with the many tests performed on the 100  $\mu\text{m}$  thick HyViSI detectors and the abundance of data exhibiting the core-halo persistence, a clear cut model for the persistence remains elusive. To truly pin down the physics behind this phenomenon, we require a more detailed knowledge of the detector layout and the ability to “turn more knobs”. For instance, changing the thickness of the detector and the dimensions of the  $\text{p}^+$  implant and gradient in the doping density at its metallurgical junction (neither of which are known because of proprietary reasons), using a metal grid under the front surface to adjust the surface potential above the  $\text{SiO}_2$  layer, etc., and observing the effect on the latent images would reveal a great deal about how and where in the detector they arise.

Despite the lack of information and knobs, based upon what is known of the detector layout and the empirical evidence presented in the preceding sections, we can present a sound, qualitative hypothesis that explains the key features. Some aspects of the model have been touched on in the previous sections. What follows is a comprehensive description that aims to address all of the components of the core-halo persistence.

### 7.5.1 Phenomenological Description

As alluded to previously, the similarity of the persistent signal induced by illumination and that brought about by forward biasing the photodiodes suggests that excess minority carriers stored in the depletion region are responsible for the latent images and persistence. In  $\text{p}^+$ - $\text{n}$  junctions used for switching applications, it is assumed that the stored minority carrier charge is dominated by holes in the bulk  $\text{n}$ -region closest to the junction when the bias direction is switched [122] from forward to reverse. We will assume the same thing here, and our assumption is supported by the fact that the persistence in HyViSIs occurs at low values of  $V_{SUB}$  when the depletion region in the PIN diodes is confined to the area near the metallurgical  $\text{p}^+$ - $\text{n}$  junction. On the slightly doped  $\text{n}$ -side of the junction (the I region), minority carrier holes are trapped near this junction when the diode is forward biased or when it is illuminated. In the former case, the holes diffuse to the region from the  $\text{p}^+$  implant and have a density profile that exponentially decays with distance into the  $\text{n}$  type bulk. In the latter, a fraction of the photo-generated holes are trapped there either before or after they drift under the influence of the electric field with a similar profile.

The role that the stored holes play in a subsequent, reverse-biased exposure of the detector is somewhat indirect. Because a large reverse current cannot flow across all depleted portions of the PIN diode, they are not immediately swept out of the diode during the reset. Instead, they add positive charge to the charge density profile, which in turn requires an additional amount of electrons to be injected into the  $\text{p}^+$  side of the junction to restore the voltage in the implant to  $V_{RESET}$ . The situation is illustrated in Figure 7.17 and explained in more detail in the figure caption. Using Gauss’s Law and integrating the electric field for a given charge distribution shows that the

number of electrons needed to compensate the trapped holes far exceeds the number of trapped holes themselves. However, the exact number will depend on how the trapped holes are distributed near the junction and how the injected electrons fill the  $p^+$  implant during reset. Since the doping profiles of the PIN diodes is not known, neither of these two distributions can be calculated. As seen in Figure 7.17, it is assumed that the trapped holes and injected electrons are exponentially distributed around the metallurgical junction.

After the reset switch is closed, as illustrated in Figure 7.18, the effect of the injected electrons is two-fold. First, electrons that diffuse to the sites of the trapped holes and recombine with them will cause a rise in signal. This occurs because of the fact that the positive charge from the trapped hole is effectively removed from the charge density after the electron recombines with it, which lowers the

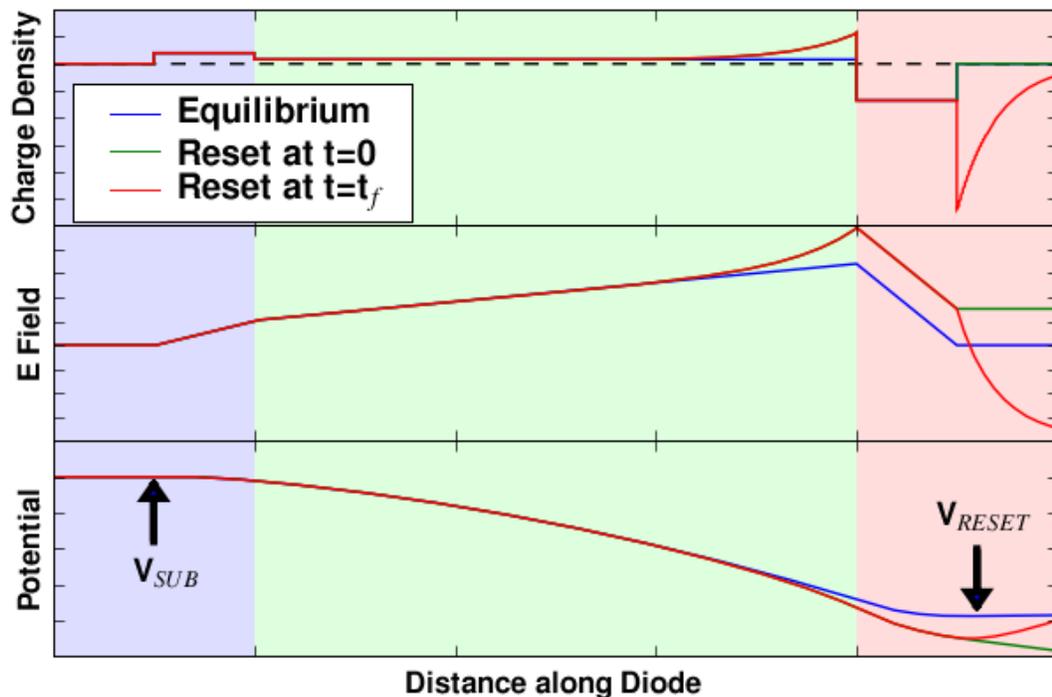


Figure 7.17: A toy model showing the charge density, electric field, and electric potential in a PIN diode in various circumstances. The electric field is obtained using Poisson's equation and integrating the field yields electric potential. The  $p^+$  region is shaded in light red, the  $n$  bulk in green, and the  $n^+$  region in blue. The diode in equilibrium, i.e. after a long series of resets in the dark, is shown by the blue curves. After illumination or forward bias, there is an excess of holes stored near the  $p^+$ - $n$  junction. At reset, if they were not compensated by injected electrons, the green curves would show the three distributions (the green curves are hidden in some regions because they lie beneath the red curves). With the injected electrons from the reset well, the potential at the  $p^+$  implant is brought to  $V_{RESET}$  and the distributions look like the ones shown by the red curves.

magnitude of electric field and decreases the potential across the PIN diode. This process, indicated by  $U_n = U_p$  in Figure 7.18, increases  $V_{node}$  and hence the signal measured, i.e.  $\Delta S > 0$ . Second, the drift and diffusion of the electrons itself can cause both an increase or decrease in signal, depending on the relationship between the direction of their motion and the direction of the electric field. This follows from the fact that the work done on an electron moving through an electric field (by the field) is the line integral of the electric field over the path taken by the electron. If the field does positive work on the electron the voltage across the diode will decrease and if it does negative work then the voltage across the diode will increase.<sup>7</sup> The field around the metallurgical junction, an approximation of which can be viewed in Figures 2.6 and 2.7, possesses a large amplitude and is directed to do positive work by pulling holes toward the  $p^+$  implant and electrons away from it. If this field is preserved and an electron diffuses or drifts from the  $p^+$  implant to the  $n$  type bulk, then  $\Delta S > 0$ . For an electron diffusing in the opposing direction,  $\Delta S < 0$ .

This bi-modality of signal change is believed to be the source of the core-halo structure. Free electrons—sourced by the reset and accumulated near the front surface in between the  $p^+$  implants—are transported between pixels via the currents  $J_{DRFT}^{n,x}$  and  $J_{DIFF}^{n,x}$ . Pixels that see a decrease in the number of electrons over time in the collecting node will have  $\Delta S > 0$  and those that see an increase will have  $\Delta S < 0$ . The exact rates of signal change will be determined by the drift, diffusion, and recombination as well as the intricate interactions of the electrons near the Si-SiO<sub>2</sub> interface and at the  $p^+$ - $n$  junctions. However, it will now be shown that one can produce many of the key features of the core-halo persistence with the first three terms alone.

### 7.5.2 Persistence Simulations

According to the model just described the dominant interactions in the core-halo persistence, which are illustrated in Figure 7.18, occur at or near the front surface of the detector and are governed by the following equations

$$\frac{\partial n}{\partial t} = \nabla \cdot (\mathbf{J}_{Diff} + \mathbf{J}_{Drft}) - U_n \quad (7.19)$$

$$\frac{\partial p}{\partial t} = -U_p. \quad (7.20)$$

No transport currents exist for the holes because they are confined to deep level trapping sites. If the doping densities of the silicon and geometries of the pixel were known, these equations could be solved on a two or three dimensional grid by choosing an appropriate set of initial conditions

<sup>7</sup>An easy way to see this without working through the math is to consider two relevant cases. In one, the electric field in the depletion region of the reverse biased PIN diode does *positive* work on an electron-hole pair generated by a photon, and this leads to a *decrease* in the voltage across the diode. In the other, the built in voltage across a PN junction is established by the electrons and holes diffusing against the electric field arising from the ionized donors left behind. The electric field does *negative* work on the diffusing carriers and this corresponds to an *increase* in the voltage across the junction.

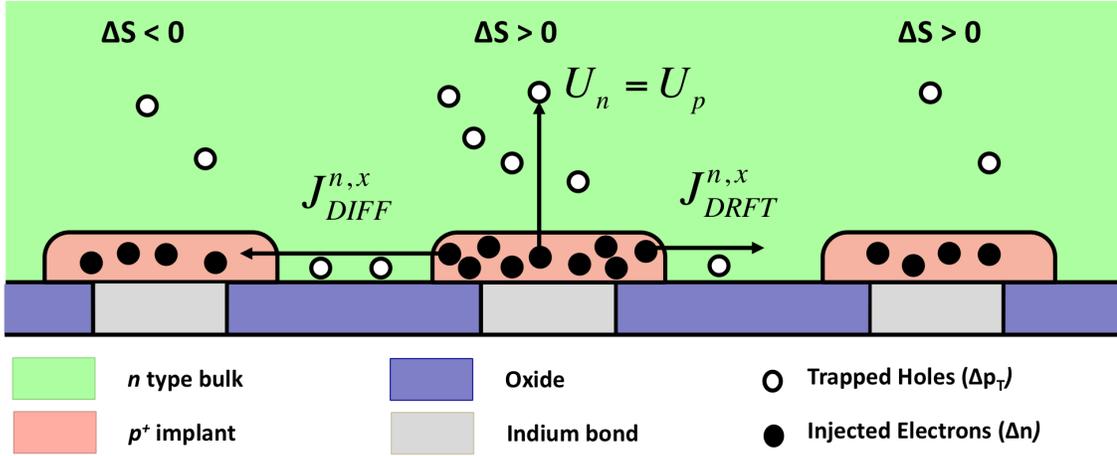


Figure 7.18: A diagram showing the diffusion and drift currents and recombination believed to be responsible for the Core-Halo persistence. The trapped holes require a surplus of injected electrons at reset to restore the pixel potential to  $V_{RESET}$ . Some of the injected electrons drift or diffuse to the trapping sites and recombine with the holes at the rate  $U_n = U_p$ , which causes an increase in the pixel signal,  $\Delta S > 0$ . Other electrons drift or diffuse away from the pixel with a current of  $J_{DRFT}^{n,x}$  or  $J_{DIFF}^{n,x}$ , respectively. Pixels that lose the electrons see a signal gain,  $\Delta S > 0$ , and those that receive them see a decrease,  $\Delta S < 0$ . Note that the plane 1/2 of a pixel length behind the plane of the drawing would have only bulk n type material, making it an avenue for the electrons to travel through if the surface is accumulated.

and using a finite difference scheme like the one outlined in Appendix C.3. Since these specifics are unknown, a more simplified approach shall be taken to model the system.

To start with, the two-dimensional doping pattern of the detector created by the p<sup>+</sup> implants is ignored and the detector is treated as being uniformly doped. This treatment is justified if the electron transport occurs primarily in the n type bulk in the “alleys” between implants. Then, assuming the core-halo persistence occurs in a sufficiently thin layer near the front surface of the detector ( $z = 0$ ) and utilizing rotational symmetry around the core center, the effect can be simulated using a quasi-one-dimensional computational scheme. In this scheme, the computational details of which are outlined in Appendix C.5, an initial distribution of excess holes,  $p(r, t = 0)$ , and electrons,  $n(r, t = 0)$ —both of which are assumed to be present after the detector is illuminated and the pixels are reset—are evolved in time in cylindrical coordinates. For simplicity, the ionized donors and equilibrium carrier distributions present in the silicon are ignored and attention is paid only to the trapped holes and injected electrons. The vertical component of electric field is also neglected so that only the radial component, which is assumed to arise solely from  $p(r, t)$  and  $n(r, t)$ , affects the evolution of the system. While the explicit currents and field are confined to the  $z = 0$  plane, the simulation effectively allows for the loss of carriers in the transverse direction through the terms  $U_n$

and  $U_p$ , making it quasi-1d (this is explained in more detail in Appendix C.5).

After establishing the mesh for the computation and the current transport equations, the only things that remain are to determine the recombination rate and the initial conditions. In this model, the trapped holes are allowed only to recombine with the injected electrons so that the recombination rates are necessarily the same  $U_n = U_p$ . A very simplified form of the Shockley-Read-Hall Equation [112] is used to describe the rate of recombination:

$$U_n = U_p = -\frac{pn}{\tau_p(n+p)}, \quad (7.21)$$

where  $\tau_p = \tau_n$  is meant to represent the lifetime of the free electrons when present amongst the trapped holes. The initial distribution of trapped holes  $p(r, t = 0)$  is based directly on the distribution of collected charge generated by the illumination source before reset, i.e. the signal measured in the previous exposure. In the case where the detector is undepleted, a piecewise function like the one in Equation 2.14 is appropriate. For the overdepleted case, the following form is used:

$$p(r, t = 0) = \begin{cases} Q_{FW} + Q_{SAT} * \exp^{-r^2/(2\sigma_{sat}^2)} & 0 < r < r_{sat} : \text{Saturated by Illumination} \\ Q_{FW} * \exp^{-(r-r_{sat})^2/(2\sigma_{psf}^2)} & r_{sat} < r < r_{max} : \text{Not Saturated} \end{cases} \quad (7.22)$$

Based upon the argument presented in the previous section, the electron distribution will have the same functional form, but the number of electrons in a given pixel will exceed the holes by some amount:  $n(r, t = 0) = \gamma p(r, t = 0)$ .  $\gamma$  is in general a nonlinear function that depends on the number of holes as well as their true arrangement along the  $z$  axis. Solving for it would involve not only a knowledge of this arrangement, but an iterative numerical procedure to solve for the electron distribution that sets  $V_{node} = V_{RESET}$ . Here it will be treated as a constant:  $\gamma \sim 4$ .

The simulation itself only provides the number of electrons and holes at a given location and time. The key assumption made in comparing the simulation to real data is that the change in  $n(r, t)$  for the pixels is directly related to the change in signal,  $\Delta S_{ADU}$ :

$$\Delta S_{ADU}(r, t) = G_{NET} \Delta n(r, t) = G_{NET} \cdot [n(r, t) - n(r, t = 0)]. \quad (7.23)$$

$G_{NET}$  is the conversion gain between  $e^-$  and ADU.

### 7.5.2.1 Simulation Results

The simulation that will be described here was intended to mimic core-halo persistence induced by the star HD53791 during a 30 second  $g$  band exposure, in which the star saturated out to a radius of  $r_{sat} = 6$  pixels. Following the exposure, the detector was blanked off and a 200 second, 500 read dark exposure was recorded in window mode with a frame time of  $t_{frame} = 0.46$  s. The illuminated exposure was used to deduce the initial conditions. The parameter  $Q_{FW}$  was computed using the

Table 7.1: Parameters for the core-halo simulation shown in Figure 7.19.

$Q_{FW}$ ( $e^-$ )	5560	$\mu_n$ ( $cm^2/s/V$ )	$3.24 \cdot 10^{-7}$
$Q_{SAT}$ ( $e^-$ )	585	$D_n$ ( $cm^2/s$ )	$1.62 \cdot 10^{-10}$
$\sigma_{sat}$ (pix)	2.15	$\tau_p$ (sec)	50
$\sigma_{psf}$ (pix)	3.70	$r_{sat}$ (pix)	6

approximate saturation of the center pixel and the relationship of  $0.027$  latent  $e^-/e^-$  noted in Section 7.2.2.2, and  $\sigma_{psf}$  was based on the FWHM of the stars in the saturating exposure.  $\tau_p$  was set at 50 seconds based upon the longest time constant,  $\tau_2$ , observed in the experiment described in Section 7.2.2.3 under the assumption that this is the time constant associated with the recombination of trapped holes and injected electrons. The other values are based largely on trial and error until a suitable match between the simulation and the real data was found.

Figure 7.19 shows the results for a simulation of the latent image in the dark exposure following the illumination; the initial conditions and other relevant variables are listed in Table 7.1. The radial profile taken at  $t = 200s$  for the simulated data shows good quantitative agreement with an

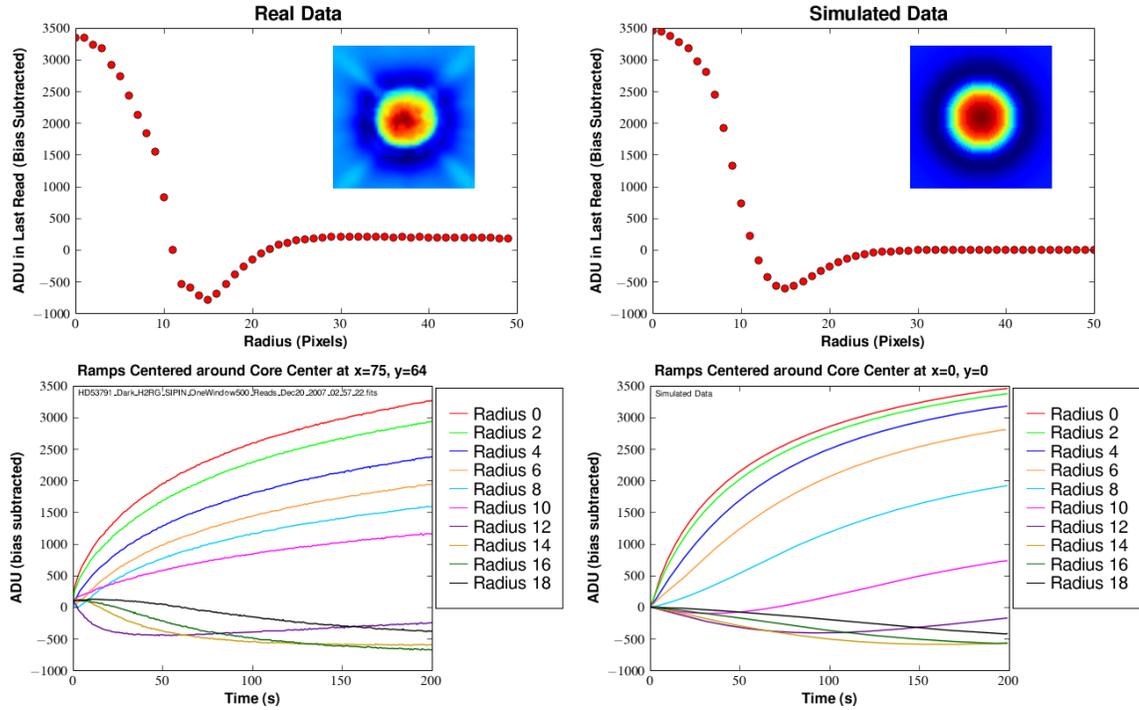


Figure 7.19: A comparison of persistence observed after a long exposure and simulated persistence.

averaged profile from the real data. The average for the real data was taken across several rows in order to avoid the non-radially symmetric diffraction spikes that show up in the latent image. One can still see some oddity in the real data near the halo minimum—it is not smoothly varying—that has to do with the diffraction pattern and the telescope tracking error. The latter can be observed in the sub-images included with the plots. The pixel ramps for the simulated data show good qualitative agreement with the real data, especially for the pixels near the core center and at the outer edge of the halo, but there are some large discrepancies for the pixels between  $R_{core} = 9$  and  $R_{min} = 15$ . In general, the real data shows that the halo pixels closest to the latent image center see a very rapid drop in counts. In fact, Figure 7.8 shows that the time constants associated with the decay of the halo are very close to those of the rise of the core. However, this is not reproduced in the simulation, where the halo pixels have a much lower rate for the decay (this can be seen most easily by comparing the purple and pink curves). The junction fields or the two-dimensional structure of p<sup>+</sup>-n implants—the diffusion and drift is being modeled as occurring only in the n type bulk, but neighboring pixels can actually be thought of as a p channel transistor where rapid carrier transfer may take place—may be responsible for the rapid decay of the halo minimum. Despite these shortcomings, the overall similarities are remarkable considering this is a simple quasi-1d simulation.

Another noteworthy point is that the diffusion coefficient and mobility obtained from the simulation are extremely low for electrons in silicon. Typical values for the mobility of electrons in silicon are  $\mu_n = 10^2 = 10^4 \text{ cm}^2/\text{V/s}$ , which is more than 10 orders of magnitude larger than the value obtained in the simulation. Small diffusion coefficients on the order of  $10^{-10} \text{ cm}^2/\text{V/s}$  are more typical for ions in SiO<sub>2</sub> [123], but the presence of ions in the HyViSI SiO<sub>2</sub> layer is extremely unlikely (this can be said for any device fabricated with modern semiconductor processing techniques in a clean environment). Further, the value of  $D_n$  from the simulation is about  $10^3$  times as large as  $\mu_n$ , which conflicts with the Einstein relation  $D_n/\mu_n = kT/q$ . Since the data was taken at 160 K,  $kT/q \sim 0.014V$ . One might therefore guess that the electric field is negligible in the simulation, but as Figure 7.20 shows, this is not the case. The final electron and hole distributions are very different depending on whether or not drift is included. The primary difference is that with drift included, the mutual repulsion of the electrons drives them away from the core at a much faster rate than with diffusion alone. This steepens the profile between  $R_{core}$  and  $R_{min}$  and creates a closer match to the data. It also aids in decreasing the decay time of the halo pixels, bringing it closer to the rise time for the halo pixels. And for long simulation times, the field actually causes some of the electrons that left the core to return there. This behavior is essential to account for the decrease in core signal for long integration times, as exemplified by Figure 7.4.

Another important point made by Figure 7.20 is that after 200 seconds, a large number of trapped holes still remain. If the detector is again reset, this distribution of holes will cause another population of electrons to be injected into the core pixels. The holes would thus serve to determine the initial conditions for a simulation of an exposure following the reset. In simulations where  $t_{final}$

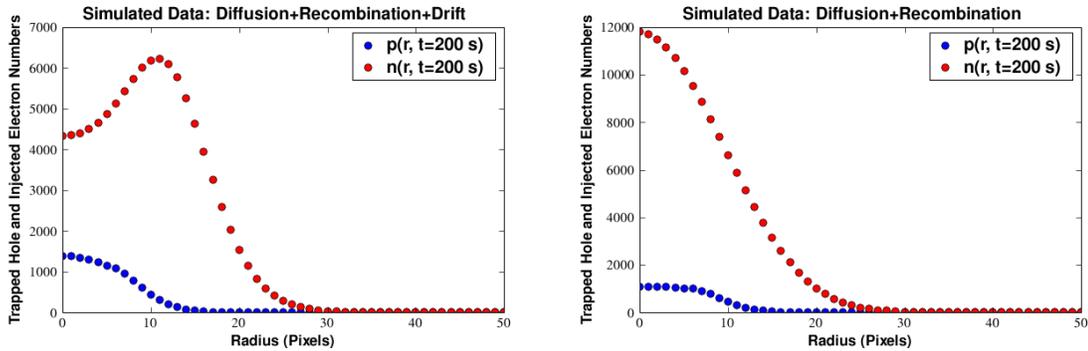


Figure 7.20: A comparison of the simulated hole and electron distributions with and without drifting the electrons. The field that creates the drift is due to the injected electrons and trapped holes.

is on the order of 10 minutes, most of the electrons migrate to the outer radii, creating a scenario where the number of holes in the core pixels exceeds the number of electrons. Since the number of electrons must be at least as great as the number of trapped holes to remove the trapped holes through recombination, this scenario explains how the persistence can last for hours.

Although not shown here, with the appropriate initial conditions, the simulations can also account for the double core-halo structure when the detector is not fully depleted. In both the over-depleted and under-depleted case, the simulations are successful in reproducing the transition of a pixel far away from the stimulated region from integrating dark current to suddenly seeing a signal change in the opposite direction. It has been noted that the dark current volcanoes also show a halo of pixels that decrease in signal, so one might hypothesize that this model could also be applied to them. However, the transitional behavior just described does not occur for the pixels far away from the center of the volcanoes, suggesting a different mechanism behind them.

### 7.5.3 Summary

To summarize, a model has been proposed to explain latent images in the HyViSI detectors, particularly the core-halo structure. The model states that a fraction of holes generated from illumination, electrical stimulus, or cosmic ray events are caught in deep level traps somewhere near the front surface of the detector. At reset, the presence of these trapped holes forces a large number of electrons to be injected by the reset transistor in order to bring the voltage at the integrating node to the reset voltage. These injected electrons either recombine with the trapped holes or migrate away from the collecting node. Pixels that see a net increase in the number of electrons at the collecting node decrease in voltage and vice versa. The persistence subsides after the fraction of traps that are populated by holes is restored to its equilibrium value.

This model accounts for many of the traits exhibited by the persistence and latent images. The

most important are:

- 1) **The core-halo structure itself.** Emission of holes from traps in the near vicinity of the  $p^+$  implant alone cannot account for the negative signal in the halo. Charge must be transferred between pixels. The drift and diffusion of the injected electrons explains both the positive signal in the core and the negative signal in the halo.
- 2) **Persistence at low  $V_{SUB}$ .** The model states that the stored charge after a stimulus is located near the metallurgical  $p^+$ -n junction at the front side of the detector. Hence, the layer of undepleted bulk for values of  $V_{SUB} < 7-8$  V does not prevent the persistence from occurring. It only changes the behavior by allowing a wide diffusion of holes to occur during the stimulus and create stored charge at larger radii.
- 3) **The double core-halo structure at low  $V_{SUB}$ .** When the detector is not fully depleted and the pixels are saturated by light, diffusion in the undepleted regions spreads the holes out to far radii and they fill trapping sites along the way. The existence of two strong gradients in the trapped hole concentration—one at the photocurrent source and one near the edge of the diffusion envelope—results in a double core-halo structure in the latent image.
- 4) **The temperature dependence of the halo width.** To produce a halo, the model requires that electrons move along the front surface. The diffusion and drift of electrons along this surface requires that it is in an accumulated state. At a fixed surface potential, the induced charge density depends very strongly on temperature through the product of the Debye length,  $L_D \propto T^{1/2}$ , and the intrinsic carrier concentration,  $n_i \propto T^{3/2} \exp(-E_g/2kT)$ , where  $E_g$  is the energy gap [124]. For low temperatures ( $T \sim 110K$ ) the persistence is dominated by recombination of the injected electrons with the trapped holes and very little transport along the front surface takes place, so the halo has a very small extent. As the temperature increases, the accumulation along the surface allows for easier transport of the injected electrons and the halo width grows accordingly.
- 5) **Persistence without saturation of pixels.** According to the model, stored charge will be present after stimulus regardless of whether or not the pixels are saturated. For signal in the range  $0 - FW$ , the amount of stored charge should be proportional to the amount of holes reaching the front side of the detector. This supports the linear relationship between the integrated photocurrent in the stimulus exposure and the latent image maximum in the one following.
- 6) **Saturation of persistence signal.** The nonlinearity and eventual saturation of the latent image maximum as a function of integrated photocurrent in a previous exposure occurs because the number of available trapping sites decreases as the photoholes accumulate. While a more sophisticated model would be required to explain the manner in which these sites

are occupied—and where exactly they are located for that matter—this model requires the persistence to saturate at some level since the number of traps is finite.

- 7) **Sensitivity of Core-Halo Pixels.** The name “desensitization” has been used to describe this phenomenon under the assumption that the pixels suffered a loss in sensitivity. However, the empirical evidence shows that both the core and halo pixels in a latent image still integrate photocurrent (if starlight is shining on them, for instance). The charge transport of injected electrons in the model does not preclude the pixels in the latent image halo from integrating signal charge, so they are not desensitized.
- 8) **Eventual decay of core pixels in long time exposures.** A large fraction of stored holes may be left behind in the core pixels after the injected electrons have migrated to the halo. The electric field created by these holes eventually pulls the electrons back into the core. Those that diffuse back into the  $p^+$  implants in the core cause a decrease in signal.
- 9) **Reappearance of latent images after long time exposures.** The left-behind holes can survive for times much greater than  $\tau_p$  if there are no electrons with which to recombine. If stored holes still exist at the core when the pixels are reset, a new swarm of electrons will be injected, giving rise to another latent image. The model can thus account for the disappearance of persistence in 200-300 seconds when regular resets are applied and the reappearance of persistence after 1 hour or more when only one or two resets are applied.
- 10) **Similarity between illumination and electrical stimulus.** Both illumination and a forward bias on the diodes creates an increase in the number of minority carrier holes in the  $n$  type bulk right around the  $p^+$ - $n$  junction. Thus, according to the model, both result in occupancy of the deep level traps. With illumination, gradients in the charge distribution result because of non-uniformity of the photon flux, and these gradients create a core-halo structure. With a forward bias all pixels are uniformly affected (ignoring pixel to pixel non-uniformity) and so no core-halo structure is formed. Instead, the persistence is dominated by the recombination of the injected electrons with the stored holes so that all pixels see an increase in signal. And since the electrons do not migrate horizontally, they stay with the trapped holes and recombine. This means that given sufficient time, the persistence will disappear independent of the number of resets performed. With the same line of reasoning, the model predicts that the time constants for the rise of the core pixel signals after illumination will be shorter than those for the rise of the pixel signals after electric stimulus.
- 11) **Persistence is not detected in reset frames.** With the reset transistor actively delivering electrons to compensate for the stored charge, the measured voltage at the integrating node stays at  $V_{RESET}$ .

## 7.6 Dealing with Persistence

Smith et al. present a number of possible ways to lower persistence in HgCdTe arrays through alterations in design or by using physical techniques such as waiting for the fast time constant portion of the decay to take place before beginning another exposure [116]. They also present methods to calibrate the persistence so that it can be accounted for and removed if present in science data [125]. Some of these strategies are applicable to SiPIN diodes, but others are not due to the fact that HyViSI detectors are operated in full depletion while the HgCdTe are per-pixel depleted detectors. The disparity in the depletion state of the two types of detector creates some fundamental differences, most notably with the time constants involved—the decay time constants in HgCdTe detectors depend on stimulus while they appear to be independent of stimulus at a given temperature in HyViSI detectors—and the lack of a core-halo structure in the HgCdTe detectors.

As discussed, the core-halo latent images in HyViSI detectors are only well fit with Equation 7.13 for a very narrow range of circumstances. And even with the simulations described in Section 7.5.2, there are significant residuals after subtracting the real data from the simulated data. Calibrating and removing the persistence from science data is thus not likely to be fruitful for these detectors. Instead, the course should be taken to prevent the persistence from occurring in the detector through design improvements or minimizing it with operational strategy.

### 7.6.1 Persistence Reduction by Design

It is very difficult to suggest design improvements without knowing the design itself. Such information is considered proprietary to Teledyne Imaging Sensors. But with or without an intimate knowledge of the architecture, what follows are some possible approaches to take.

#### 7.6.1.1 Frontside Passivation Treatments

The theory that the persistent charge is due to stored minority holes in the vicinity of the  $p^+n$  junction would imply that attention should be paid to this region. It is likely that the majority of the trapping sites are located at the interface between the  $\text{SiO}_2$  passivation layer and the Si bulk since these interfaces are notorious for having large trap densities. Some manufacturers aim to reduce the number of these trapping sites by treating the surface with special techniques during fabrication. In fact, at least three of the devices tested during the course of this thesis work were made with dissimilar surface treatments (Bai, private communication). Since the persistence was equally strong in all of them, alleviating the problem in this way may not be a viable option.

#### 7.6.1.2 Controllable Potential Metal Grid

The use of a metal grid below the surface passivation (see Figure 1.7) that can be held at a controllable potential is one option to consider. Raytheon Vision Systems uses this approach [38], which

was successful in drastically reducing persistence in InSb arrays [42]. Building a HyViSI with the metal grid would, at the very least, allow one to confirm whether or not the stored charge is located near the front surface passivation. In the best case scenario, placing the surface at a higher potential than the  $p^+$  implants could be used to effectively create true potential wells and prevent holes from reaching the trapping sites altogether. The downside of this grid is that it increases the capacitance of the pixels, thereby reducing the sensitivity, and may result in yield issues.

### 7.6.1.3 Anti-blooming Diode

The first generation H4RG had an anti-blooming diode in the multiplexer pixel. As evidenced by the data collected with H4RG-10-007, this diode prevented column bleeding, output crosstalk, and blooming. It did not eliminate persistence, but set a threshold on the maximum observed persistence signal that was well below the other HyViSI detectors. This is in agreement with the model presented since the number of holes available for trapping does not increase after the blooming diode starts to bleed them off. The problem with the anti-blooming diode implemented is that it creates unacceptable leakage currents, which led TIS to remove it from their next generation H4RG [88]. Because of the benefits it presents, though, future consideration should be given to reimplementing the anti-blooming diode.

### 7.6.1.4 Capacitive TransImpedance Amplifier (CTIA)

In the SFD architecture, photo-charge accumulates in the collecting  $p^+$  implant, which allows charge to be trapped in the vicinity of the  $p^+$ - $n$  junction as it builds up. In the CTIA architecture, the  $p^+$  side of the diode is connected to the integrating node of a feedback circuit so that the dominant accumulation takes place on a feedback capacitor in the ROIC. In theory, this should all but eliminate persistence since the ability of the minority carriers to accumulate around the front side of the detector is significantly hindered. However, the CTIA has many disadvantages for astronomy when compared to the SFD. It consumes more power, has a higher readout noise, and requires a large pixel footprint in the multiplexer, which limits the possibility of reducing pixel size [116, 23].

## 7.6.2 Reduction After Design

So far, no operational techniques have been successful at eliminating persistence in the sensors where it has been observed. With the current design of the detector it appears that the best one can do is minimize its strength and duration using the following strategies.

### 7.6.2.1 Short Exposure Times

Because the latent image maximum increases with increasing fluence, the time taken for the persistence signal to reach the read noise level does as well. Short exposure times will limit the fluence

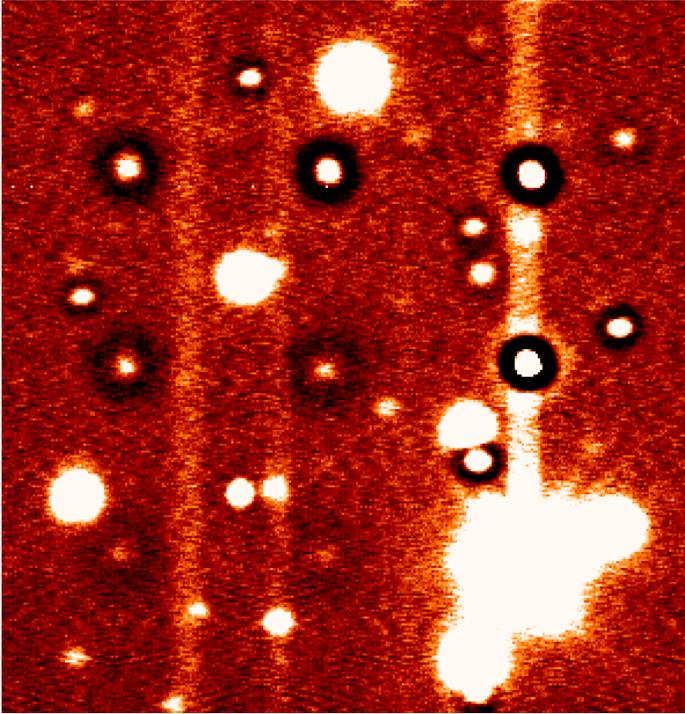


Figure 7.21: An example of a poorly chosen  $3 \times 3$  dither sequence that shows latent images in all of the previous 8 dither locations. The exposure time was 21s and the time between exposures was 16s so the dither period was  $T_{Dither} = 333s$ . This is shorter than the time it takes for the persistence for the brightest star in the image to fall below the read noise level, so the latent image from the first dither is still detectable at about the  $1\sigma$  level. When the dithers are combined, the core-halo images will show up in several locations (see the mosaic in Figure 6.20). An interesting overlap between the light from a star and the halo of a latent image is also shown.

seen by the pixels so that the latent images fall below the read noise more quickly. The latent images will still decay with the same time constant, of course, so care must be taken when planning a dither sequence. If the ratio of the time period for the dither sequence,  $T_{Dither}$ , to the decay time of the persistence is too short, the result will look like Figure 7.21, in which latent images show up at all of the previous dither locations. In this instance, when combining the dithered exposures, the mean or median of the nine pixel values used to estimate the flux at a given location on the sky will be thrown off by up to 4 values tainted by persistence. One solution is to simply do more dither locations with the same exposure time, thereby increasing the number of samples used to estimate the flux. Another solution is to use a slightly longer exposure time. For instance, if the exposure time in the  $3 \times 3$  sequence in Figure 7.21 was increased to 120 s, at least four out of the nine locations would not have latent images above the read noise level. It is always possible to mask the core-halo structures in software when combining the dithers, but this will result in increased noise for certain spatial locations.

### 7.6.2.2 Frequent Resets

It may seem that frequent resets and short exposures imply the same thing. This is true if the detector is operated in full frame mode since two separate exposures means a reset of all pixels between the two. However, if the detector is operated in window mode, a reset sequence only resets

the pixels within the window; the other pixels on the detector continue to integrate charge.

For window mode applications such as telescope guiding, whenever window data is not required, the detector should be switched to full frame mode so that all the pixels on the array can be reset. The reason for this is that if intermittent full frame resets are not performed when the detector is operated in window mode, bright stars may saturate pixels outside the window and generate persistence. This persistence will cause havoc when the detector is later read in full frame mode or a new guide window is selected that happens to fall on a previously saturated set of pixels. The full array may be reset line by line to ensure stability in signal after the reset, but this comes at the expense of a large reset time for the full frame. The reset time can be reduced dramatically if the detector is switched to full frame, global reset mode and then a global reset is performed (the detector manual states this reset should be greater than 10  $\mu\text{s}$  in order to fully reset the pixels and empirical evidence shows it should be less than  $\sim 100$  ms to prevent strong nonlinearities). However, after the global reset, when the detector is switched back to window mode, a line by line reset of the window should be employed to prevent nonlinearities in the early reads of the window.

The empirical data shows that a reset should be performed at least every 120-200 seconds at  $T=160$  K to clear the trapped charge in the shortest possible time. With the right initial conditions, simulations show that after this time period, trapped holes still exist near the core with no electrons available for recombination, so another reset must be applied to provide them. When resets are only performed every 20-30 minutes, the persistence can endure for 1-2 hours. So if long exposures are a must, 300-400 second long reset periods should be issued between them to prevent the latent images from showing up in each of the long exposures.

### 7.6.2.3 Lower Temperature and Higher $V_{SUB}$

Persistence is still observed at the lowest temperature,  $T=100\text{K}$ , tested for this thesis work and the highest backside voltage,  $V_{SUB} = 40\text{V}$ . But because the halo width decreases when the temperature is lowered and the PSF is decreased when  $V_{SUB}$  is raised, both of these changes result in a smaller number of pixels affected by the latent images.

### 7.6.2.4 Reset of Saturated Regions with Guide Windows

One possible way to prevent the buildup of stored charge via large fluences is to reset saturated regions during long exposures using the guide mode of the HxRG multiplexers. This requires either prior knowledge of the relevant pixel locations or “smart firmware” that automatically detects saturated regions and programs the HxRG serial register with the appropriate window coordinates and then resets these regions.<sup>8</sup> Complex clocking patterns and instruction sets would be thus be involved. The signal change of the rows and the columns of the window, which were discussed in

<sup>8</sup>Detection of the saturated regions could also potentially be done on the DAQ, but a dedicated set of instructions in the control electronics microcontroller or FPGA would not impose on the flow of science data between the two systems.

Section 6.1.4.3 and are shown most clearly in Figure 5.11, would create an additional complication when the data is reduced.