L3 CCD Simulation

T.R. Marsh

October 28, 2005

1 Introduction

The purpose of this document is to detail the mathematics of the statistics of the avalanche readouts of L3 CCDs with the purpose of investigating how they can be simulated. Only a little of the stuff here is original, but some of the background is not laid out in an easily accessible form in the papers I have read, so I thought it would help to have everything in one place. It is rather mathematical and if you prefer not to plough through all the (many) equations, then you may just want to look at the figures and take a look at section 6 which summarises the main points.

2 The Model

In the avalanche gain section of the L3 CCD, high clock voltage causes there to be a significant probability that an electron will produce another electron when moving from stage to stage. In common with others, I will call this probability p. In practice it is fairly small, of order 0.015. However there are many such multiplication stages and so the overall gain can be large.

A schematic of the process is shown in Fig. 1. The key thing to note is that at any stage if there are, say, n electrons, then each of them independently goes through the same multiplication process for the next stage. i.e. it is not the case that if there are 100 electrons that there is a p chance at the next stage that there will be 200 (instead it will be p^{100}).

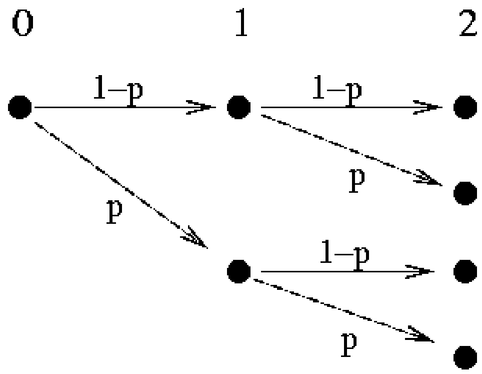

Figure 1: Schematic of the multiplication process, starting from 1 electron on the left at stage 0. At stage 1, a second electron is created with probability p. Each of the two possible electrons at stage 2 then independently can go through the same multiplication, so that as many as four electrons may be present by stage 2.

3 Statistics

Key properties of interest are the mean gain as a result of a given number r stages of this process, the variance of the gain which is important because it is a measure of the noise added by the avalanche process, and, in most detail of all, the probability distribution of the gain. This was worked out by Matsuo et al (1985) although they do refer to difficult-to-access texts for several results; I will repeat their analysis here which is based upon a general theory of “branching processes” which goes under the name Galston-Watson branching process, which was first considered in the context of extinctions of populations in the 19th century. I do this fairly fully so that this document is self-contained.

3.1 Fundamentals

First I will assume that we start from one electron at stage 0. I am interested in the distribution at stage r, which I will associate with a random variable called N_r. The probabilities we want are $P(N_r = n)$. For instance these give us the mean gain after r stages g_r from

$$g_r = E(N_r) = \sum_{n=0}^{\infty} P(N_r = n)n, \quad (1)$$

where I use the standard notation $E(X)$ to mean the "expected value" of a random variable X. However, it turns out that it is not possible to obtain a simple expression for the probabilities except in special cases, and we have to adopt a more sophisticated approach. To do so we will consider two functions known as the "moment generating function" μ (MGF) and the "probability generating function" π (PGF) which for a random variable N are defined by

$$\mu_N(t) = E(e^{tN}) = \sum_{n=0}^{\infty} P(N = n)e^{tn}, \quad (2)$$

and

$$\pi_N(t) = E(t^N) = \sum_{n=0}^{\infty} P(N = n)t^n, \quad (3)$$

These two functions are closely related to each other as it is easily seen that

$$\pi_N(t) = \mu_N(\ln t). \quad (4)$$

If one takes the derivative of the MGF then

$$\frac{d\mu_N(t)}{dt} = E(Ne^{tN}), \quad (5)$$

and therefore

$$\left.\frac{d\mu_N(t)}{dt}\right|_{t=0} = E(N), \quad (6)$$

which is the first moment (mean) of N that we want. It is then obvious that

$$E(N^p) = \left.\frac{d^p\mu(t)}{dt^p}\right|_{t=0}, \quad (7)$$

and hence the term "moment generating function". The idea is that sometimes the moment generating function is easier to handle than the probability distribution and so it is possible to obtain moments when once cannot obtain (easily) the probabilities themselves. Similarly, from the definition of the PGF one can show that

$$P_N(n) = \frac{1}{n!}\left.\frac{d^n\pi_N(t)}{dt^n}\right|_{t=0}, \quad (8)$$

and hence this is the "probability generating function".

The MGF and PGF are standard functions in statistics. A final result which we use below is that if one has two independent random variables, N and M then

$$\begin{aligned}\pi_{N+M}(t) &= E(t^{N+M}), && (9)\\ &= E(t^N)E(t^M), && (10)\\ &= \pi_N(t)\pi_M(t). && (11)\end{aligned}$$

The first to the second line uses the property that for two independent random variables X and Y, $E(XY) = E(X)E(Y)$. An identical result applies to MGFs.

3.2 The PGF and MGF for avalanche multipliers

Consider the PGF at stage $r + 1$ of the avalanche multiplier

$$\begin{aligned}\pi_{r+1}(t) &= E(t^{N_{r+1}}) && (12)\\ &= \sum_k E(t^{N_{r+1}}|N_r = k)P(N_r = k), && (13)\end{aligned}$$

where we have written $\mu_{N_{r+1}} = \mu_{r+1}$ etc for short. The last line introduces an expectation conditional on a particular value of $N_r = k$ for the previous stage. This is relatively easy to calculate because if one knows that there were k electrons at stage r, then the distribution for stage $r + 1$ is simply the combined result of k single electron multiplications, $(N_1)_1 + (N_1)_2 + (N_1)_3 + \ldots + (N_1)_k$ where each element is distributed as N_1, the distribution after 1 stage with a $1 - p$ chance of 1 electron and a p chance of 2 electrons. This has corresponding PGF (dropping the subscript 1 because we will use this special case a fair bit)

$$\pi(t) = (1 - p)t + pt^2, \quad (14)$$

and thus the PGF for the combination of k such variables is, by Eq. 11, $= \pi^k(t)$, and therefore from Eq. 13 we have

$$\begin{aligned}\pi_{r+1}(t) &= \sum_k P(N_r = k)\pi^k(t), && (15)\\ &= \pi_r(\pi(t)). && (16)\end{aligned}$$

We can immediately write from this that

$$\pi_2(t) = \pi(\pi(t)), \tag{17}$$
$$\pi_3(t) = \pi(\pi(\pi(t))), \tag{18}$$
$$\pi_4(t) = \pi(\pi(\pi(\pi(t)))), \tag{19}$$

etc, and also therefore that

$$\pi_{r+1}(t) = \pi(\pi_r(t)) \tag{20}$$
$$= (1-p)\pi_r(t) + p\pi_r^2(t), \tag{21}$$

the last line following from Eq. 14. A similar relation applies to the MGF:

$$\mu_{r+1}(t) = (1-p)\mu_r(t) + p\mu_r^2(t). \tag{22}$$

Eqs 21 and 22 are recurrence relations that can be applied to deduce the PGF and MGF at stage r given that $\pi_0(t) = t$ and $\mu_0(t) = e^t$. They correspond to equation 4 from Matsuo et al (1985).

3.3 The mean and variance

We are now in a position to deduce the mean gain, because if we take the derivative of Eq. 22 and set $t = 0$ then by Eq. 6 and using the fact that $\mu(t = 0) = 1$ for any MGF, we have

$$g_{r+1} = (1+p)g_r. \tag{23}$$

Given that $g_0 = 1$ then we have

$$g_r = (1+p)^r, \tag{24}$$

which is what one would have hoped. To derive the variance we want the second moment, and therefore the second derivative of Eq. 22 which after setting $t = 0$ leads to

$$E(N_{r+1}^2) = (1+p)E(N_r^2) + 2pg_r^2. \tag{25}$$

Using $\sigma^2 = E(X^2) - E(X)^2$, leads to

$$\sigma_{r+1}^2 = (1+p)\sigma_r^2 + p(1-p)(1+p)^{2r}, \tag{26}$$

where we have substituted the relation for g_r. Starting from $\sigma_0 = 0$, one can show that this recurrence relation is satisfied by

$$\sigma_r^2 = \frac{1-p}{1+p}\left(g_r^2 - g_r\right). \tag{27}$$

As $r \to \infty$ and therefore $g_r \to \infty$, $\sigma_r/g_r \to \sqrt{(1-p)/(1+p)} \sim 1$ for small p. This is the root cause of the extra noise added by the avalanche gain section: the gain can be large but it is also uncertain by a similar order of magnitude. As $p \to 1$, $\sigma_r \to 0$, reflecting the fact that if multiplication is certain, then the gain just becomes 2^r with no spread. Unfortunately, this is not a realistic case in practice. The uncertainty in the gain in practice is why L3 CCDs operated in a strict proportional mode effectively have half the QE of a standard CCD.

3.4 The probability distribution

According to Eq. 8, to get the probability distribution we must take the n-th derivative of Eq. 21 and divide by $n!$. This leads to the recurrence relation

$$P_{r+1}(n) = (1-p)P_r(n) + p\sum_{k=0}^{n} P_r(k)P_r(n-k), \tag{28}$$

where $P_r(n)$ is a short-hand form of $P(N_r = k)$. This can be computed given that $P_r(0) = 0$ and $P_0(1) = 1$, however, it does not lead to analytically tractable expressions, and one must resort to computation. Computing it directly can take a long time. For example, on my desktop, taking $p = 0.015$ and $r = 591$, then $g = 6629$. It turns out that the probability is approximately $\exp -x/g$, and so in an exact computation one would be interested in numbers up to a few times g. For instance, computing up to gains of 30,000 would not be unreasonable in this case. On my workstation this takes 600 seconds, and since it is dominated by the sum in the above equation, it scales as the maximum number squared. Luckily, since the sum is a convolution, FFTs can help out, and the same computation with FFTs only takes 22 seconds which is an acceptable overhead for the start of a data simulation for instance. FFTs do have one disadvantage in this case which is that round-off errors limit how low one can go in probability, but computing in double precision gives a reliable dynamic range of ~ 14 orders of magnitude which should be fine.

3.5 Clock-Induced Charges

During clocking it is possible for electrons to be generated without the aid of incident photons. Normally these are buried in readout noise but in an L3CCD they will be amplified and therefore be significant. Any generated in the parallel shifts will undergo precisely the same amplification as photon-generated events, and will be indistinguishable from them. Those generated in the avalanche register however will have a different distribution because if say a CIC appears in the last 10 section of the register, they will not be amplified much, and so we expect a skew towards much lower numbers of electrons. There seems to be some uncertainty over which will dominate but I have been told that it is the in register events that are most important because of the high voltages (Basden, prov. comm.). Therefore from now on I will only discuss these latter events.

The probability distribution of pure CICs can be considered in the case of no input electron. Then at each step of the avalanche register let there we a small probability p_{CIC} of an electron being generated. If this happens on step k of the total r, then it will be amplified over the next $r-k$ steps leading to a contribution with distribution P_{r-k}. The overall contribution from step k will in fact be

$$P'_k(n) = (1 - p_{CIC})\delta_{0n} + p_{CIC}P_{r-k}(n) \qquad (29)$$

and the final total distribution will be the convolution of all such distributions for $k = 1$ to r. Convolution once again suggests FFTs, and the calculation of the final distribution can in fact be nicely tacked on to the FFT computation of P_r at little extra computational cost. One of the nice features of these computations is that there is no "downscatter" in that the probability of a number of counts n at any stage depends only upon earlier probabilities for 0 up to n and not on any higher values. This allows an exact computation up to some preset limit.

Fig. 2 shows the results of computations of CIC distributions for several different assumed values for p_{CIC}. The point about no "downscatter" is visible in the lack of "edge" effects on the right-hand side of this figure which is exactly set at the highest value considered.

In photon counting mode, one sets a threshold above which a pixels is counted as having a photon. To avoid loss of sensitivity this must be set as low as possible, but not so low that many spurious events are generated by readout noise. For example, in the case shown if Fig. 2 a level of 100 would result in only a small loss of counts (about 2%) while being much greater than a readout noise of, say, 10 electrons RMS. However, the figure shows that it may instead be CICs that really set the threshold, depending upon p_{CIC}. For instance if $p_{CIC} = 0.0032$, then a threshold of 100 would see $\sim 60\%$ of pixels with spurious "photons" due to CIC. Since in photon counting mode one wants to keep the object signal well below a mean rate of 1 photon/pixel, this would be a disaster. Raising the threshold to 1000 would reduce the rate to $\sim 30\%$ but also lose 1 in 6 or so of the real counts.

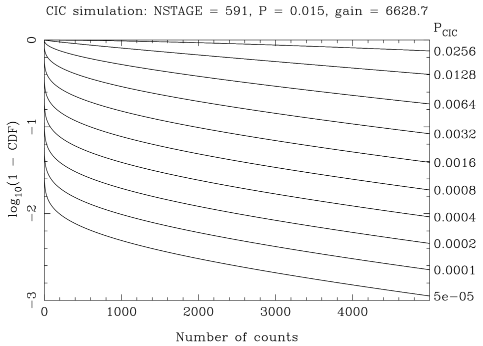

Figure 2: The logarithm of the probability of obtaining a CIC $> x$ is plotted as a function of x for several different CIC probabilities, marked on the right, but with other parameters held fixed as indicated at the top of the plot. this shows for instance that for $p_{CIC} = 0.0008$, slightly fewer than 1 in 10 pixels will suffer a CIC with more than 1000 counts.

For the case shown, a value for p_{CIC} in excess of 10^{-4} or so will be bad news. As I understand it, the probability of CICs is a function of the voltages in the sense that increasing voltages imply increasing p_{CIC} as well as increasing gain. It may be that the increase in gain outstrips p_{CIC}, on the other hand we may find that there is some optimum gain with respect to CICs and readout noise.

At present there seems much uncertainty over a reasonable value for p_{CIC} and whether CICs in the serial register (as modelled here) dominate over parallel CICs. It may depend upon the nature of the clock waveforms. It will be important to be able to characterise CICs well in real devices. For instance by clocking out the serial register without any parallel clocks (if possible). My hope is that like readout noise in the old days of CCDs, CICs are a problem that will get better with time.

3.5.1 The mean output from CICs

Since the output probability distribution of CICs is the convolution of the P'_k distributions discussed above, then the MGF of the ouput is given by

$$\mu_{CIC} = \Pi_{k=0}^{r-1} \left(1 - p_{CIC} + p_{CIC}\mu_k\right). \quad (30)$$

Taking the derivative in order to compute the mean gives

$$\frac{d\mu_{CIC}}{dt} = \sum_{k=0}^{r-1} p_{CIC} \frac{d\mu_k}{dt} \Pi_{l\neq k} \left(1 - p_{CIC} + p_{CIC}\mu_k\right). \quad (31)$$

Setting $t = 0$, and remembering that $\mu(0) = 1$ gives the mean CIC-only output to be

$$\begin{aligned} g_{CIC} &= \sum_{k=0}^{r-1} p_{CIC} g_k && (32) \\ &= p_{CIC} \sum_{k=0}^{r-1} (1+p)^k && (33) \\ &= \frac{p_{CIC}}{p}(g-1), && (34) \end{aligned}$$

where $g = (1 + p)^r$ as before. This nicely shows that it is important that $p_{CIC} \ll p$.

4 Simulating L3 CCDs

We need the probability distributions discussed above to simulate the action of L3 CCDs. This can be done by computing the exact probabilities as outlined in the previous section and then searching a look-up table. Basden et al (2003) found that for small p and large r that the distribution for a single electron input ignoring CICs was fairly well matched by an exponential distribution of the form

$$P(x) = \frac{1}{g} e^{-x/g}. \quad (35)$$

I find that the following function

$$P(n) = \frac{1}{g-1} \left(\frac{g-1}{g}\right)^n, \quad (36)$$

which has a mean of g and a variance of $g^2 - g$ does a rather better job, especially at small g. The fact that this is not quite the exact variance of Eq. 27 is because this probability distribution is only approximate. In reality it tends to be too large for $n \ll g$ and $n \gg g$ and too small for $n \sim g$. Still, the approximation is still pretty close to the mark as far as I can determine and makes a quick way to generate the distribution to compute

$$N = \text{int}\left(1 - \frac{\ln X}{\ln g/(g-1)}\right), \quad (37)$$

where X is a uniform random number between 0 and 1.

All the above holds for a single input electron. If m electrons are input then each will be independently amplified and so the resulting probability distribution will be the convolution of m of the above probability distributions. Basden et al (2003) give the following approximation for their exponential distribution:

$$Q_n(m) = \frac{m^{n-1} \exp(-m/g)}{g^n (n-1)!}, \quad (38)$$

which gives the probability of an output of m electrons given an input of n. I use the letter Q because this probility represents the chance of an output of m given an input of n, as opposed to the previous usage of the chance of an output of n at stage r. This equation would be an approximation, even if their single electron probability was correct, but it is a pretty good one in many cases as I will show in the next section.

For my improved single electron distribution, Eq. 36, I find the following exact expression for a convolution of n such distributions:

$$Q_n(m) = \frac{(m-1)!}{(m-n)!} \frac{1}{(g-1)^n (n-1)!} \left(\frac{g-1}{g}\right)^m, \quad (39)$$

which applies so long as $m \geq n$ and $n > 0$. For instance setting $n = 1$ immediately returns the single electron input case. This is still approximate since the single electron input is approximate, and at high gains at least it is indistinguishable from Basden et al's (2003) approximation, i.e. it fits the real distribution as well (or as poorly) as Basden et al. It may be possible to find a similar analytic approximation for CICs, although they are certainly not simple exponentials.

One could repeatedly instead sample the 1 electron distribution and add the results; whether this

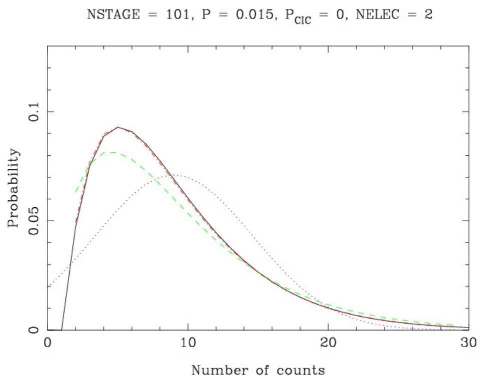

Figure 6: The probability distribution for 2 input electrons in a low gain case. This brings out the advantage of the revised approximation Eq. 39 (red dot-dash line) compared to Eq. 38 (green dashed line).

is feasible is probably a matter of computational time. The CICs can always be added in from a single lookup.

I do not yet know the practicality in terms of time of implementing look-up tables once they have been calculated. I expect them to be slow, but not unreasonably so. For example, one could look-up from a table of 65000 values in about 16 "binary chops", which does not seem too bad.

5 Example Distributions

In this section I just present some example distributions in Figs. 3 to 6. Key points are

1. In-register CIC events have very extended tails (Fig. 3)
2. Basden et al's (2003) approximation for multiple input electrons is better than a gaussian of the correct mean and variance up to quite high input numbers. (Fig. 5)

6 Executive summary

These are the main points to take from this work:

1. The statistics of the gain produced by L3 CCDs are somewhat complicated, but they can be calculated in a reasonable amount of time, so I hope that accurate data simulations will be possible.
2. At high gain the distribution given a single electron at the start of the avalanche regsister is close to exponential as in Eq. 35 and as shown in the top-left of Fig. 4.
3. An approximation which works somewhat better at lower gains is given in Eq. 36.
4. Eqs 35 and 36 have multi-electron equivalents: Eqs 38 and 39. The equations don't do well when the multiplication probability is large. Examples are shown in Figs 4 to 6.
5. Clock-induced charges (CICs) are potentially serious, but be very careful with any figures you see quoted for them because there is a lot of folklore/hogwash floating around about them.
6. CICs generated on the chip are indistinguishable from normally generated events, but might show up as a gradient in the parallel direction, with higher numbers of events the further a given pixels has had to travel.
7. CICs generated in the serial register can have an extremely skewed distribution with many very small events, but also a long tail extending to large counts (Fig. 3).
8. In photon counting mode at least CICs simply add extra background, the crucial but unknown quantity being just how much do they add.
9. Characterisation of CICs is important. It would be nice to be able to clock out the serial register without parallel shifts for instance because then one can determine the importance of events generated in the serial register alone.

I have written various programs to compute the distributions discussed here and can easily generate many such plots.

7 References

Basden, A.G., Haniff, C., Mackay, C., 2003, MNRAS, 345, 985

Matsuo, K., Teich, M.C., Salej, B.E.A, 1985, IEEE, Transaction on electron devices, Vol. ED-32, No. 12, 1223

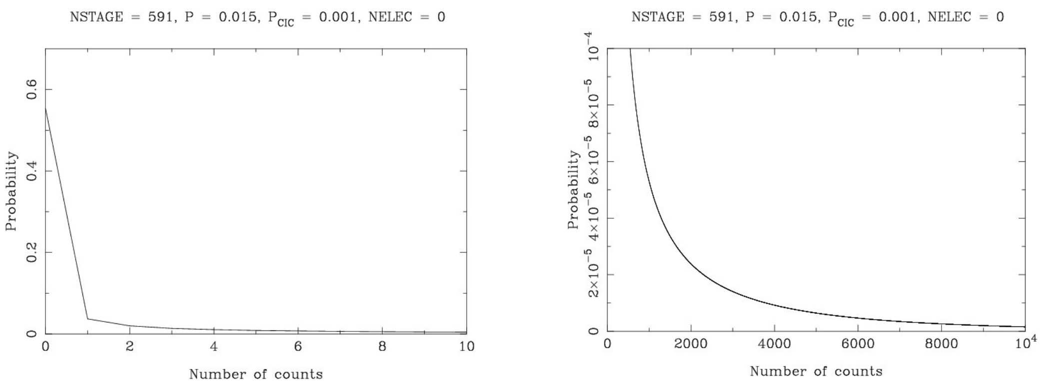

Figure 3: The probability distribution for 0 input electrons, i.e. pure CIC events. I show a zoomed in one on the left to show the significant chance in this case of no CIC events at all, and a larger scale one on the right to show the very extended tail of the CIC events, which is this case have a mean value of 442 despite having a 55% chance of being zero.

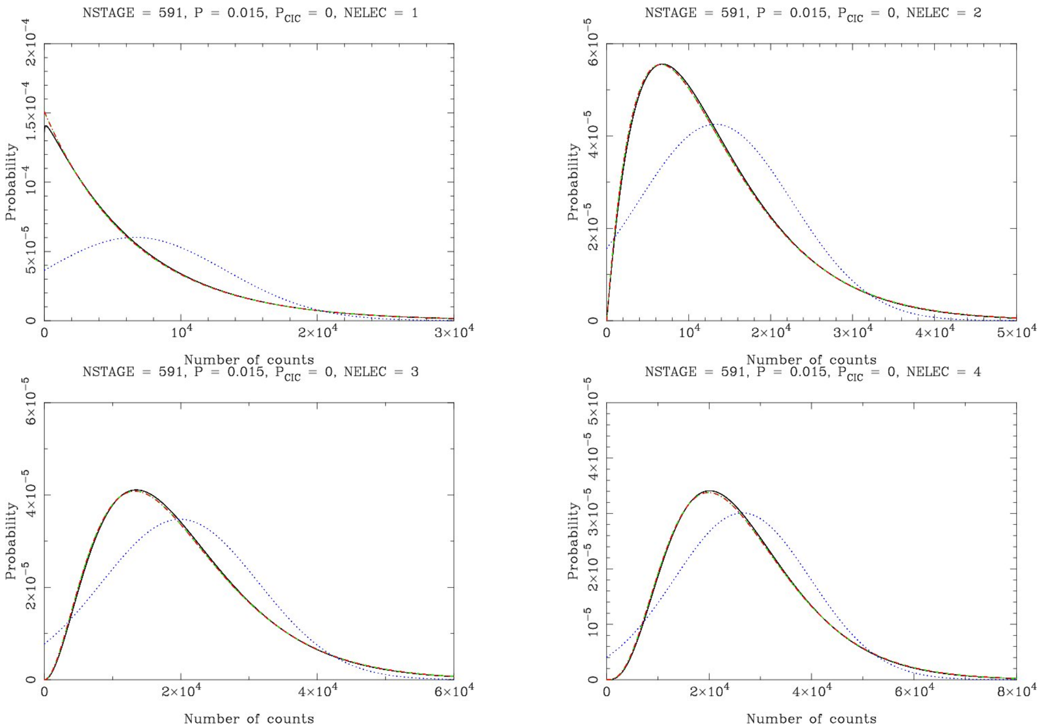

Figure 4: The probability distributions for 1, 2, 3 and 4 input electrons with no CICs. The green dashed line shows Basden et al (2003)'s approximation, the red dash-dotted line shows my approximation, while the blue dotted line shows a gaussian of the same mean and variance (which the distribution should tend towards by the central limit theorem). Note that the scales change on each plot.

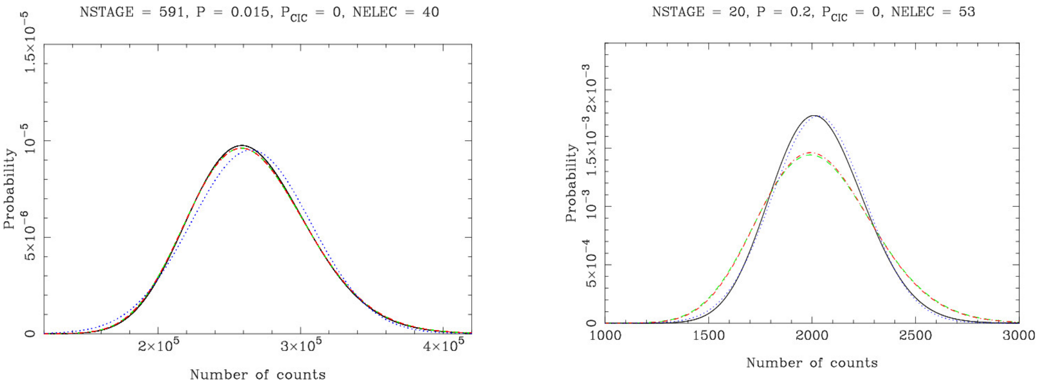

Figure 5: The probability distribution for 40 input electrons with no CICs. The various lines are as in Fig. 4. The distribution remains significantly skewed and so the approximate expressions remain better than a gaussian even for this relatively large number of input electrons. However the right-panel shows that gaussians can be better especially when p is large.