arXiv:physics/0301002 v1 1 Jan 2003

Random sampling of an AC source: A tool to teach probabilistic observations

Arvind,* Paramdeep Singh Chandi,† and R. C. Singh‡
Department of Physics, Guru Nanak Dev University, Amritsar 143005 India

D. Indumathi§ and R. Shankar¶
Institute of Mathematical Sciences, CIT Campus Taramani, Chennai 600113 India

An undergraduate level experiment is described to demonstrate the role of probabilistic observations in physics. A capacitor and a DC voltmeter are used to randomly sample an AC voltage source. The resulting probability distribution is analyzed to extract information about the AC source. Different characteristic probability distributions arising from various AC waveforms are calculated and experimentally measured. The reconstruction of the AC waveform is demonstrated from the measured probability distribution under certain restricted circumstances. The results are also compared with a simulated data sample. We propose this as a pedagogical tool to teach probabilistic measurements and their manipulations.

PACS numbers: 01.50.Pa,01.55.+b

I. INTRODUCTION

Probability is fundamental to physics in more ways than one. Probabilistic errors can never be avoided in experimental observations, individual particles and their initial conditions cannot be tracked in classical physics, and quantum mechanics, which is the best available description of nature, is intrinsically probabilistic. While the basic concepts of probability can be introduced nicely through coin tossing and probability boards[1], they remain setups in the realm of statistics without a direct connection to the physics laboratory. The pedagogy of probability for physics students has received attention, with many proposals of statistics-oriented experiments[2,3,4,5] and theoretical expositions[6,7,8,9], providing interesting insights. A more physical example of radioactive decay, a natural random (quantum) process has also been used to teach and demonstrate probabilistic ideas[10,11,12,13]. However, a typical physics laboratory requires more experiments involving probability distributions and their manipulation. This paper is an effort in this direction where we describe a simple gadget to study the manipulation of probability distributions.

Imagine inserting the terminals of a DC voltmeter into the AC mains outlet. We expect it to display zero because the DC meter will respond to the average voltage which in this case is zero. The DC voltmeter is not designed to be sensitive to changes of voltage which occur at the frequency (50-60Hz) of a typical AC source. Therefore, the only information we can get from such a measurement is the average voltage.

How does one measure the instantaneous voltage? We need to "store" this value for long enough that a DC voltmeter will be able to read it. One way to do this is to connect a capacitor across the AC source. The capacitor will get charged; the instantaneous voltage across it will determine the instantaneous charge on the capacitor plates. When the capacitor is disconnected from the circuit, the charge and hence the voltage, on the capacitor remains. This can be seen by joining the two terminals of the capacitor, whereby a spark is produced. This voltage can be measured by a high impedance DC voltmeter. The DC voltage measured across the capacitor is then the instantaneous AC voltage of the original source. This DC voltage is the key variable measured in our experiment.

The experiment can be repeated and a different voltage will be obtained each time! If the observation is repeated many times, it is indeed a random sampling of the AC voltage source. This randomly sampled voltage data can be used to construct the probability distribution of the voltage. What information about the AC source is contained in this probability distribution? We will see that the probability distribution depends upon the type of AC waveform used. A triangular wave for example, will give rise to a very different probability distribution as compared to a sine wave. Furthermore, under certain restricted circumstances we can reconstruct the waveform from the voltage probability distribution. We explicitly demonstrate such a reconstruction for the case of a sine wave.

The results presented here are an instructive demonstration of the role of probabilistic analysis in physics experiments. The apparatus is simple and cheap. Elementary C-programs running on an ordinary PC are sufficient to accomplish the data analysis. The data analysis can also be performed using a graph paper, pencil and a pocket calculator. Computer analysis is not essential but is instructive and opens up possibilities of playing around with various parameters.

The experiment can be introduced in a physics laboratory course at several different levels. At the lowest level the data collected by the procedure described in section II can be used to demonstrate the zero mean, maximum and minimum values and the RMS voltage of an AC source. At the next level of the undergraduate physics laboratory, the analysis of Sections II, III, and IV can be used to calculate the voltage probability distribution for different parameter values and to reconstruct the corresponding waveforms. At a more advanced

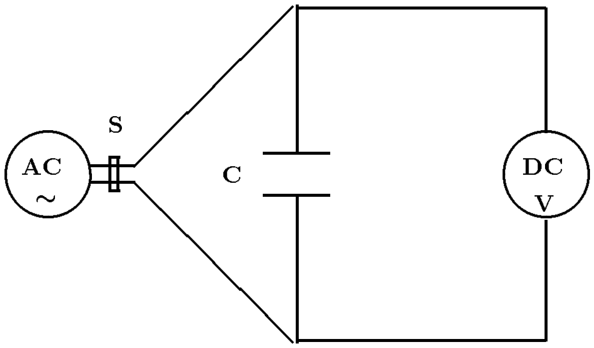

FIG. 1: The circuit diagram of the experiment. An AC source is connected to a capacitor C through a switch S which when pulled disconnects both the terminals of the source from the capacitor. The capacitor voltage is continuously monitored through a DC voltmeter V. When the source is connected the DC meter shows zero voltage and when the source is disconnected the DC meter shows a random voltage which decays as the capacitor begins to discharge through the voltmeter.

level, the statistical analysis of section V and simulations can be carried out with the help of computer programs (available in Numerical Recipes[14]) to bring out the quantitative statistical aspects of the experiment.

The material in this paper is arranged as follows. In section II we describe the experimental apparatus. Section III provides a theoretical analysis of probability distributions arising from a random sampling of voltages and waveform reconstruction from such a probability distribution. Section IV describes the experimental measurement with AC waveform and the data analysis. In Section V we compare the results of our experiments with data obtained from a simulation. The C-program used for the data analysis in Section IV is provided in the Appendix . Section VI contains a short discussion and conclusions.

II. EXPERIMENTAL SETUP

The experimental setup consists of a capacitor, a voltmeter, an AC source and a double pole switch. The voltmeter is connected across the capacitor which is in turn connected to the source through the switch. The switch when pressed disconnects the source from the capacitor. At this instant the capacitor begins to discharge through the voltmeter. Upon pressing the switch a second time, the source is reconnected to the capacitor. The voltmeter remains connected across the capacitor throughout. A good digital voltmeter with high internal resistance and small capacitance should be used. The circuit diagram is shown in Figure 1.

This arrangement is sufficient to sample the distribution of the voltage developed across the capacitor. The measurement proceeds as follows: the switch is kept in place to ensure that the DC voltmeter shows zero voltage. This happens because the AC is oscillating too rapidly for the DC meter to be able to follow the voltage. Then the switch is pressed to disconnect the AC source and at this stage the voltmeter shows the instantaneous DC voltage across the capacitor. The maximum value shown on the meter is recorded. This voltage is our main observation. The voltage across the capacitor will decay slowly as it discharges through the voltmeter. We are not interested in this decay. The switch is pressed a second time to reconnect the AC source. This completes one measurement cycle. To repeat the observation, at some stage the switch is pressed again and the maximum voltage developed across the capacitor is recorded. The experiment is repeated several times and a list of voltages is generated. This is our basic data set from which we want to draw our conclusions.

It is useful to have two students recording the data; one presses the switch and the other records the maximum voltage on the voltmeter. In some voltmeters the voltage first rises and then begins to fall. To a good approximation we take the maximum voltage to be the instantaneous voltage across the capacitor. An ideal voltmeter will not have this problem; however, no instrument is ideal so there is always a finite measurement time over which the voltage across the voltmeter builds up from zero. To accurately measure the voltage across the capacitor is an interesting exercise in itself; the best way is to actually measure the charge accumulated on the capacitor using a sensitive device like a ballistic galvanometer. It will be interesting to develop newer instruments which can be used in undergraduate laboratories and which can measure charge to a good accuracy. However, for our experiment such precise measurement of voltage across the capacitor is not required. We now turn to the theoretical analysis of random sampling and probability distributions.

III. THEORY

Consider a time-dependent observable quantity, say a voltage $f(t)$. If we measure this voltage N random times in an interval, $0 < t < T$, we can determine the distribution function $n(V)$, the number of times the measurement of f results in a value between V and $V+\Delta V$. We denote the corresponding probability distribution of values of f by $P(V)$,

$$P(V)\Delta V \equiv \frac{n(V)}{N} \tag{1}$$

Given $f(t)$, what is $P(V)$? Consider a measurement of f being done between t and $t+\Delta t$. The measured value V, will be between $f(t)$ and $f(t+\Delta t)$. i.e., between $f(t)$ and $f(t) + \frac{df(t)}{dt}\Delta t$. Let $t_i,\ i = 1, 2, ...M$, be the times

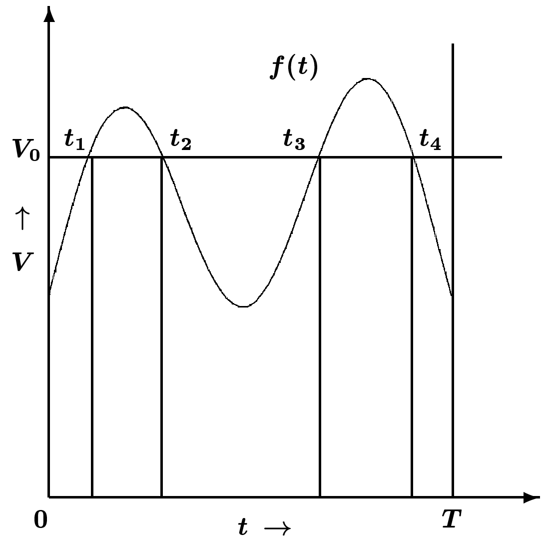

FIG. 2: Example of a typical function $f(t)$. Contribution to $P(V)$ near V_0 will come from four intervals of time in this case because $f(t)$ hits the value V_0 at times t_1, t_2, t_3 and t_4.

at which the voltage is equal to V, i.e., all the solutions of the equation,

$$f(t_i) = V \tag{2}$$

If $P_t(t)$ is the probability of the measurement being done at time t, then the probability of the outcome of a measurement being between V and ΔV is,

$$P(V)\Delta V = \sum_{i=1}^{M} P_t(t_i)\Delta t_i \tag{3}$$

where,

$$\Delta t_i = \Delta V \left| \frac{df(t_i)}{dt} \right|^{-1} \tag{4}$$

We can always invert $f(t)$ in the neighborhood of each t_i. So let $t = g_i(V)$, $t \approx t_i$. Furthermore, let us assume that the random times of measurement are uniformly distributed so that $P_t(t) = 1/T$. We then have,

$$P(V)\Delta V = \frac{1}{T}\left(\sum_{i=1}^{M} \left| \frac{dg_i(V)}{dV} \right| \right) \Delta V \tag{5}$$

Hence,

$$P(V) = \frac{1}{T} \sum_{i=1}^{M} \left| \frac{dg_i(V)}{dV} \right| \tag{6}$$

A. Examples

1. *Triangular Wave*

$$\begin{aligned} f(t) &= -V_0\left(1 - \frac{4t}{T}\right), \quad 0 < t < T/2 \\ &= V_0\left(3 - \frac{4t}{T}\right), \quad T/2 < t < T \end{aligned} \tag{7}$$

For $-V_0 \leq V \leq V_0$, every V occurs twice at the times,

$$\begin{aligned} t_1 &= \frac{T}{4}\left(1 + \frac{V}{V_0}\right) \\ t_2 &= \frac{T}{4}\left(3 - \frac{V}{V_0}\right) \end{aligned} \tag{8}$$

$g_1(V)$ and $g_2(V)$ are given by the RHS of the above equations. We can now use the formula in Eqn. (6) to get,

$$P(V) = \frac{1}{2V_0} \tag{9}$$

Independent of V. Basically, $f(t)$ is spending equal times at all voltages between $-V_0$ and V_0 and thus all voltages in this range are equally likely.

2. *Sawtooth Wave*

$$f(t) = -V_0\left(1 - \frac{2t}{T}\right) \tag{10}$$

Here, every voltage between $-V_0$ and V_0 occurs exactly once at,

$$t_1 = \frac{T}{2}\left(1 + \frac{V}{V_0}\right) \tag{11}$$

Applying Eqn. (6) as before yields,

$$P(V) = \frac{1}{2V_0} \tag{12}$$

again, independent of V. In fact the probability distributions for the triangular and sawtooth waveforms are exactly the same. In both cases, f spends equal times at all voltages between $-V_0$ and V_0.

3. *Sinusoidal Wave*

$$f(t) = V_0 \sin\left(\frac{2\pi}{T}t + \phi\right) \tag{13}$$

Each value of the voltage occurs twice,

$$t_1 = \frac{T}{2\pi}\left(\sin^{-1}\left(\frac{V}{V_0}\right) - \phi\right)$$
$$t_2 = \pi - \frac{T}{2\pi}\left(\sin^{-1}\left(\frac{V}{V_0}\right) - \phi\right) \quad (14)$$

Applying Eqn. (6) gives,

$$P(V) = \frac{1}{\pi}\frac{1}{\sqrt{V_0^2 - V^2}} \quad (15)$$

In Sections IV and V we show the comparison of this calculation with the probability distribution computed from the experimental data using a sinusoidal waveform.

B. Wave form reconstruction

Having determined the probability distribution $P(V)$, the question now is, given $P(V)$, what can we say about $f(t)$? In general, it is not possible to reconstruct $f(t)$ from $P(V)$ since many functions can have the same probability distribution of values. However, as we will see, with some additional information about the function, it is possible to reconstruct it.

We first consider the case when $f(t)$ is a one-to-one and hence invertible function. We denote the inverse of f by $g \equiv f^{-1}$, so that $t = g(V)$. Let $P_t(t)$ be the probability that a measurement is done between t and $t + dt$, then,

$$P(V)dV = P_t(t)dt \quad (16)$$

where t is the time when the voltage is equal to V. From now, we will restrict ourselves to uniform distributions for the random measurements, i.e.,

$$P_t(t) = \frac{1}{T} \quad (17)$$

We also have

$$dt = \left|\frac{dg}{dV}dV\right|,$$

hence,

$$P(V)dV = \frac{1}{T}\left|\frac{dg}{dV}\right|dV \quad (18)$$

We now need some additional information about $f(t)$. If $f(t)$ is one-to-one, then it is monotonic. Assume that it is monotonically increasing. We can then integrate the above equation to get $g(V)$,

$$g(V) = T\int_{V_0}^{V} dV' P(V') \quad (19)$$

where $V_0 = f(0)$ and $g(V)$ can then be inverted to get $f(t)$.

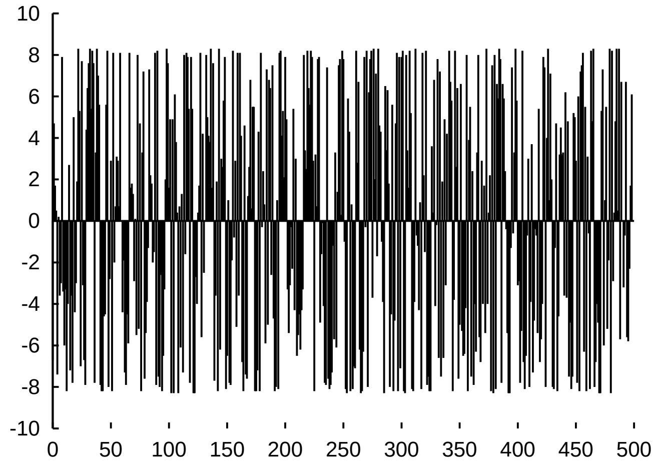

FIG. 3: Plot of 500 voltages plotted with the measurement number appearing along the x-axis. The AC source used here had a peak voltage of 8.6V.

Another situation where $f(t)$ can be reconstructed is if we are given that it is a periodic function with period T_p, has exactly one minimum (and hence one maximum) in a period and is symmetric about its maximum and minimum. A sinusoidal waveform which we use for our experiment is an example of such a function. In this case, we know that it monotonically increases for half the period and monotonically decreases for the other half of the period. Furthermore, if we are sampling over times large compared to the time period, then we randomly sample *all* the times in a period. So we can replace T by the half-period $\frac{1}{2}T_p$ in all the above formulae, reconstruct $f(t)$ for the half period and then using the symmetry, recover the function for the full period. We can get to the wave shape from the probability distribution in this case; however, we are unable to find the signal frequency.

IV. RESULTS

We now turn to the experimental results and their analysis. For good statistics, a large number of voltages should be recorded, although interesting results begin to emerge with a sample size as low as 100. The results in this section pertain to a sample size of 500 data points with a 50Hz AC source having a peak voltage of 8.6V. The 500 raw data points (used in the analysis) are shown in Figure 3. The raw data clearly look random and it is not possible to draw any meaningful conclusions by mere visual inspection.

The first step in the analysis is to choose a certain number of bins and find the probability distribution of voltage. This is achieved by the C program given in Appendix . In Figure 4 we show the graph corresponding to the probability distribution for the voltages divided into 101 uniform bins.

Furthermore, since the function in this case satisfies

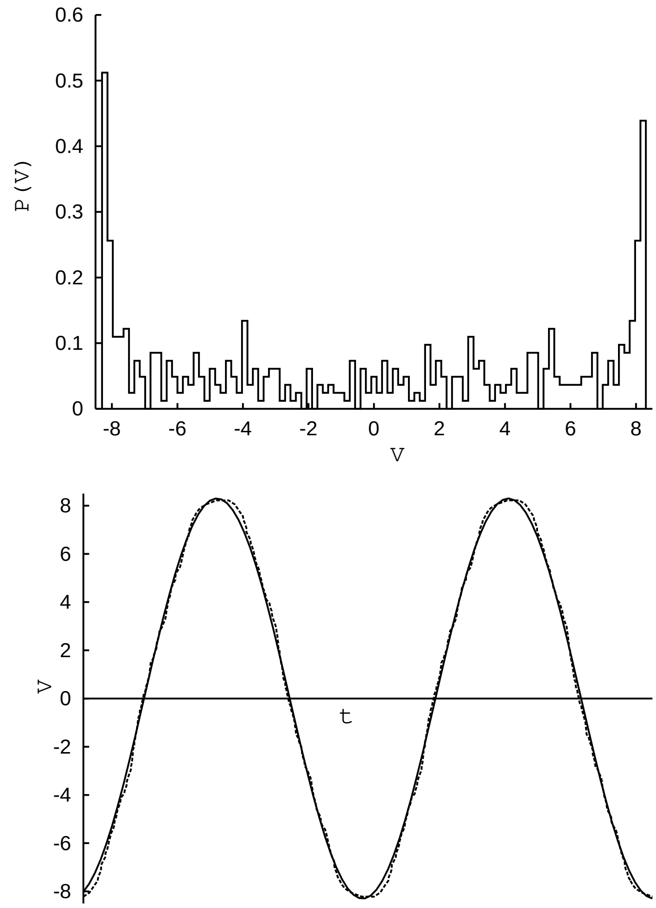

FIG. 4: Probability distribution and reconstructed waveform from 500 data points using 101 voltage bins. In the upper graph we show the probability distribution which is calculated by dividing the fraction of voltages belonging to a bin by the width of the bin. In the lower graph the dotted curve corresponds to the reconstructed voltage as a function of time in arbitrary units (no frequency information is recoverable). The solid curve is the actual sine curve provided for comparison.

the conditions for reconstruction, we use the formula given in Eqn. 19 and numerical integration (carried out by the second part of the C program given in Appendix) for waveform reconstruction and display the results in the lower half of Figure 4. The expected sinusoidal waveform emerges. The actual sine curve is also provided as a solid curve in the same graph for comparison. If we start with a larger raw data set we can improve the quality of reconstruction and also reduce the statistical fluctuation in the probability distribution plot. A detailed statistical analysis is presented in Section V. We observe here that *the probability distribution reveals the characteristics of the waveform which are not at all obvious from the raw data.*

We note here that there is no way of estimating the frequency of the signal. The sampling times, being random, do not give us any time-scale information; hence the reconstruction gives us only the shape of the signal. For example, two sine waves with different frequencies and same amplitudes will give the same probability distribution and hence the same reconstructed waveforms.

It is instructive to play with the parameters of the experiment. Different frequency and amplitude for the AC can be tried and the experiment can be repeated with different waveforms like triangular or saw tooth and others. One can change the total number of points in the raw data and see how the statistics improves by increasing the sample points, which will be taken up in the next section. For the data analysis the number of bins is the only crucial parameter and it is interesting to see how the quality of analysis changes as we change the bin size. As a demonstration, we re-analyze the same data described above by varying the number of bins to 11 and 1001. One observes (see Figs. 5 and 6) that 11 bins a give much less accurate probability distribution and waveform; however going to 1001 bins does not help much compared to 101 bins.

The 101 bins are able to capture more information available in the data compared to 11 bins, leading to the improved quality of the result. Since the raw data has only 500 points, there is a limit to which we can improve the quality by increasing the bin number. If we want to improve the results further we must increase the number of raw data points and not the bin number. This is a general principle of experimental observations wherein the accuracy of the result is determined by the least accurate part of the observation and analysis.

V. STATISTICAL ANALYSIS AND NUMERICAL SIMULATION

As seen in the preceding section, the data when analysed fit well into a sinusoidal curve, reproducing the original voltage form. We now analyse the goodness of this fit and deviations from the expected (theoretical) values.

When a sinusoidal waveform is sampled N times randomly, the voltage probability distribution of Eqn. (15) results in the frequency distribution of events

$$n(V) = NP(V)\mathrm{d}V \tag{20}$$

and the accumulated frequency of events upto a voltage V obtained on integration is

$$N_V = \int_0^V NP(V)\mathrm{d}V = \frac{N}{\pi}\sin^{-1} V/V_0 \ . \tag{21}$$

For discrete bins of size ΔV, the integration is replaced by a sum over bins, $N_V = \sum n(V) = \sum NP(V)\Delta V$. In other words, by a cumulative process of adding the frequencies in bins, starting from the $V = 0$ bin to the $V = V$ bin, we recover the sine (actually sine-inverse) form.

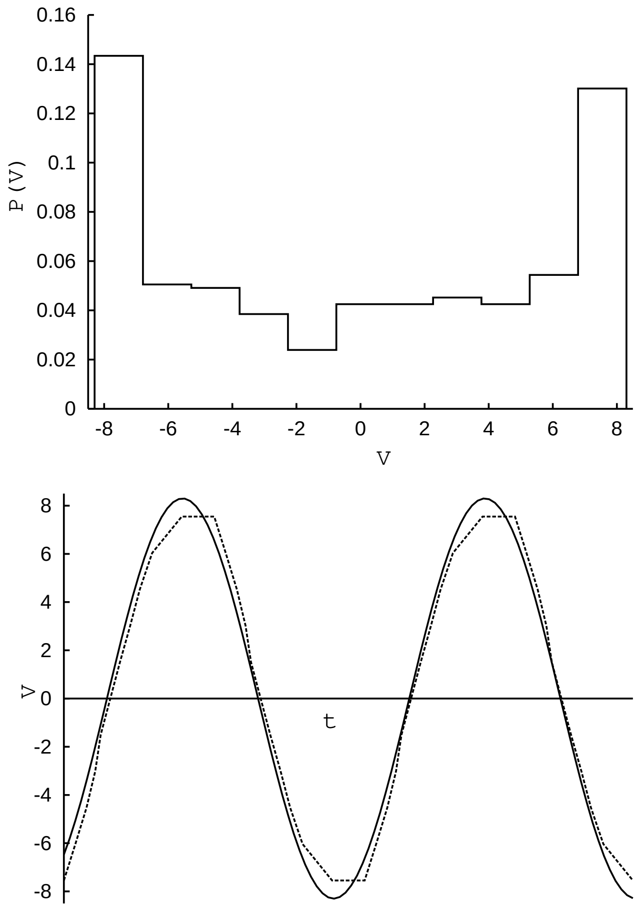

FIG. 5: Probability distribution and reconstructed waveform for the same 500 point data set but with only 11 voltage bins. The plots clearly indicate that there is a loss of information if we use too few bins for the data analysis.

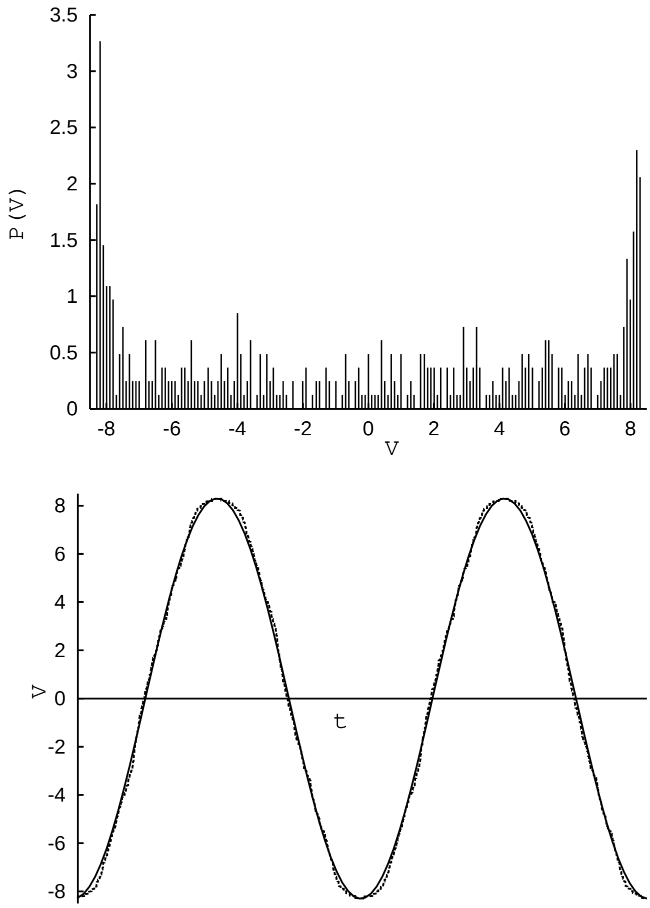

FIG. 6: Probability distribution and reconstructed waveform for the same data set with 1001 voltage bins. There is hardly any improvement from the 101 bin analysis indicating that there is no advantage in increasing the bin size beyond a point.

The N data are binned into m bins; assuming that each sample is a random independent event, the statistical error for the frequency in each bin can be taken to be $\sigma_0 = \sqrt{N/m}$. Thus the error on N_V is $\sigma_j = \sqrt{j}\sigma_0$, where there are j bins from $V = 0$ to $V = V$. With this error, the accumulated frequency N_V is fitted to the form in Eqn. (21), with V_0 as the free parameters, by a standard chi-squared minimisation procedure:

$$\chi^2 = \sum_j \frac{(N_V^{\rm data}(j) - N_V(j, V_0))^2}{\sigma_j^2} \tag{22}$$

The result of the fitting procedure for the set of 500 sample data from a transformer stepped down to peak voltage (a) 8.6 V, and (b) 17.1 V, is shown in Table I.

Note that due to switching losses, the largest value of sampled voltage in the first case was 8.3 V and in the latter, 16.4 V. Clearly, switching losses are larger at larger values of the peak voltage and the χ^2 values

V_0 (Actual)	V_0 (Measured)	V_0 (Fitted)	χ^2/dof
8.6	8.3	8.4	58/90
17.1	16.4	16.4	105/90

TABLE I: Fits to 500 samples of data of an AC Voltage with peak voltage $V_{\rm max}$, binned into 91 voltage bins.

indicate that the recovery of the waveform is truer at lower peak voltages.

We now turn our attention to a numerical simulation. Two sets of data were simulated (both with $V_0 = 8.3$V), one with 500 sample events (as in the actual experimental data) and one with twenty times as much data, by sampling a sine waveform randomly. The simulated data were binned and analysed as above. The larger data set was scaled suitably for comparison with the smaller sample/experimental data.

The error on the frequency in each bin is again $\sqrt{N/m}$.

The total frequency, $\sum_j N_V(j) = N \pm \sqrt{N} = N(1 \pm 1/\sqrt{N})$. Hence, though the error increases as $\sqrt{N}$, the *fluctuations* on the accumulated frequency decrease as $1/\sqrt{N}$; therefore, we expect the fits from the 10,000 sample set to be about 5 times $(\sqrt{20})$ smoother than those from the 500 sample set. We show the corresponding frequency vs. bin voltage histogram in Fig. 7. The histograms I, II refer to the original data set and the simulated 500-sample data. It is seen that they are very similar in appearance. The histogram III, shifted by a factor of 50 for clarity, is from the 10,000 data set and clearly shows much smaller fluctuations. Correspondingly, as shown in Table II, the χ^2 for the fits to the accumulated frequency distribution N_V are much better for III than for I or II.

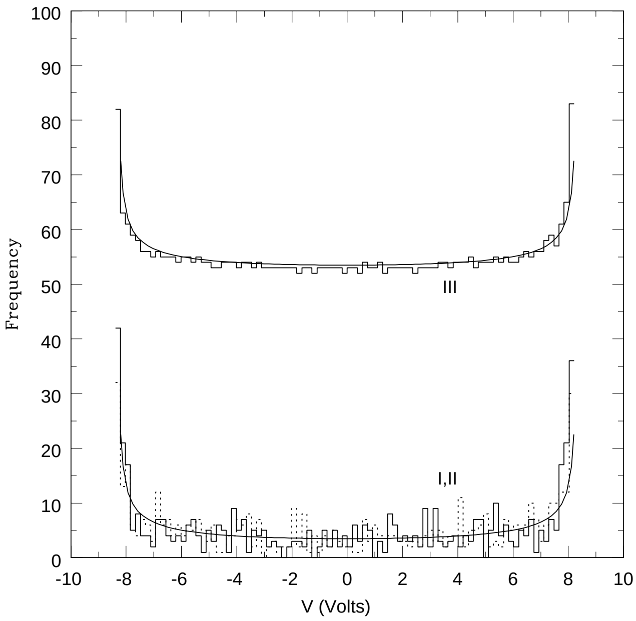

FIG. 7: Frequency of occurrence of voltage in 91 bins for a peak voltage $V_{\rm peak} = 8.6$V. Histograms I (solid lines), II (dotted lines) correspond to the experimental data and simulations with 500 samples respectively; histogram III (shifted by 50 along the y-axis for clarity) is for 10,000 samples. The solid lines correspond to the theoretical frequency distribution as given by Eqn. (20)

The accumulated frequency for the data/simulated data are shown as a function of the voltage in Fig. 8 for the three cases. (The results for III have been scaled down (by 20) to match the overall normalisation for the other two cases.) While fluctuations in the simulated data for 500 samples (Case II) are similar to the experimental data, the fluctuations for the large data sample (Case III) are very small and the corresponding distribution is very smooth. The resulting χ^2 is therefore much smaller in this case, as Table II shows. Furthermore, the goodness-of-fit is much better for II than for I, although they correspond to similar sample sizes. This may reflect the fact that Set I actually samples a waveform with peak voltage $V_{\rm peak} = 8.6$ V with voltage losses at the time of measurement due to switching; these losses may be complicated functions of V. No such losses are modeled in our analysis.

Data Set	V_0	χ^2/dof
I	8.4	58/90
II	8.3	15/90
III	8.3	2.6/90

TABLE II: Fits to 500 samples of data of an AC Voltage with peak voltage V_0, binned into 91 voltage bins. I: experimental data set, II: simulated data, III: simulated 10,000 data set, scaled to 500 samples.

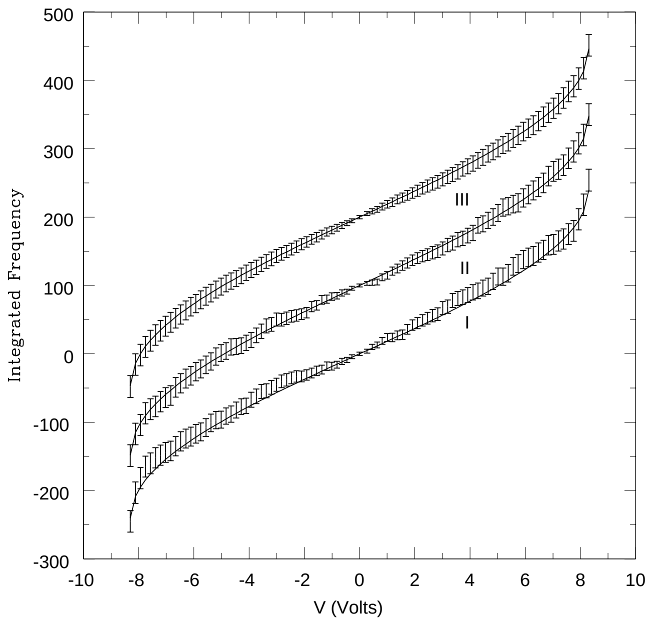

FIG. 8: Integrated frequency, N_V, shown as a function of voltage V, in 91 bins, with errorbars. The solid lines are the best fits to the data, I: experimental data with $V_0 = 8.3V$, generated by sampling AC voltage from a transformer with a peak voltage of 8.6 V. II: simulated data with same V_0 and number of sample points (500) as data, III: simulated data with 10,000 sample points. Set II (III) has been shifted by 100 (200) along the y-axis for clarity.

In summary, data with smaller voltages at the transformer give better fits to the original sine waveform than data with larger voltages. As expected, the quality of fit improves with amount of data.

VI. CONCLUDING REMARKS

A simple experiment was constructed to randomly sample an AC voltage source. The crux of the experiment involved charging a capacitor from an AC source, whose instantaneous voltage is then measured by switching its connection to a DC voltmeter at a random time. The resulting data were analysed to recover information about the AC source. While the frequency of the source could not be determined, the peak voltage and the *shape* of the original waveform could be accurately found. The procedure involved in inverting the data to recover this information was tested through numerical simulations and statistical analysis. The experiment, along with the analysis, can be effectively introduced into a physics laboratory course at the primary or advanced level. The effects of changing the switching device (to limit losses due to sparking, etc.), the peak voltage of the AC source, and the voltage measuring device, can be studied at various levels of complexity, depending on the ability and inclination of the reader.

Acknowledgments

Arvind thanks National Science Foundation for financial support through Grant Nos. 9900755 and 0139974.

APPENDIX: C PROGRAM

We give here the C program used for the data analysis. The input to the program is a data file ‘input.dat’ which should have a single column containing the voltages measured in the experiment. The program scans the file, finds the data attributes (number of data points, the maximum value in the data, the average voltage, etc.) and writes them in the file ‘cap1.out’. It then divides the voltage range into equal sized odd number of bins. The bin number is to be specified on the screen and is read as the variable `bin_nu`. The binned data is written into the file ‘cap2.out’ with the first column containing the bin center and the second the probability of occurrence of voltage in that bin. The second part of the program carries out the waveform reconstruction for the periodic signal and the result is written in the file ‘cap3.out’ where the first column contains the ‘scaleless’ time variable and the second column the voltage reconstructed for that time.

```
#include "stdio.h"
#include "math.h"
main()
{
const int bin_max=5000;    /* max array size */
const float epsilon=0.001; /* voltage range extension(end points) */
int bin_nu;                /* no. of bins to be used (must be odd)*/
int bin[bin_max];          /* array of bins */
int j,i,k;                 /* integers to be used in loops */
int data_max;              /* data points in input.dat */
int bin1_nu;
float voltage[data_max];/* array of voltages read from input.dat*/
float max_voltage;      /* maximum voltage */
float average;          /* average voltage */
float bin_nuf,bin_width,sum[bin_max];
                        /* bin no. as float, bin width and sum */
float pbin[bin_max];    /* probability bin */
FILE *fp0,*fp1,*fp2,*fp3;
fp0=fopen("input.dat","r"); /* Input file */
fp1=fopen("cap1.out","w");  /* Output file 1 */
fp2=fopen("cap2.out","w");  /* Output file 2 */
fp3=fopen("cap3.out","w");  /* Output file 3 */
printf("Input the number of bins to be used (odd number)\n");
scanf("%d",&bin_nu);
bin_nuf=bin_nu;
data_max=0;             /* Initialization */
max_voltage=0;          /* Initialization */
average=0;              /* Initialization */
for(i=0;i<bin_max;i++)  /* Initialization */
bin[i]=0;
j=1;
i=0;
while(i!=EOF)          /* Reading data from ‘input.dat’ */
```

```
{
fscanf(fp0,"%f",&voltage[j]);
i=getc(fp0);
if (i=='\n')
{
j++;
data_max++;
}
}
for(i=1;i<=data_max;i++)
{
average=average+voltage[i];
if(fabs(voltage[i]) > max_voltage)
max_voltage=voltage[i];
}
bin_width = 2*(max_voltage+epsilon)/bin_nuf; /* Computed bin width */
average   = average/data_max;                /* Average Voltage */
for(i=1;i<=data_max;i++)                     /* Filling the bins */
{
for(j=-(bin_nu-1)/2;j<=(bin_nu-1)/2;j++)
{
if(voltage[i] >=j*bin_width-bin_width/2 &&
   voltage[i] < j*bin_width+bin_width/2)
bin[j+(bin_nu-1)/2]++;
}
}
```

Calculating the probabilities for bins

```
for(j=0;j<bin_nu;j++)
{
pbin[j]=(1.0)*bin[j]/data_max;
}
```

Some basic facts about the data are computed and written in a file 'cap1.out'

```
fprintf(fp1,"Number of Voltages Scanned =  %i\n",data_max);
fprintf(fp1,"Maximum Voltage            =  %.3f\n",max_voltage);
fprintf(fp1,"Number of Bins             =  %i\n",bin_nu);
fprintf(fp1,"Average Voltage            =  %.3f\n",average);
fprintf(fp1,"Size of each bin           =  %.3f\n",bin_width);
```

Probabilities for each bin written into the output file 'cap2.out' with first column being the center of the bin and the second column the probability for finding the voltage in that bin.

```
fprintf(fp2,"Bin Center   Probability\n");
for(j=0;j<bin_nu;j++)
fprintf(fp2,"%+8.3f       %-1.4f\n",
                      (j-(bin_nuf)/2+0.5)*bin_width,pbin[j]);
```

INTEGRATION OF THE DATA

This part of the program integrates the data and reconstructs the waveform assuming that $f(t)$ can be reconstructed from the values of $f(t)$ in the interval $[0, T/2)$ (where T is the period of $f(t)$) in the following way: $f(T/2+t) = f(T/2-t) \quad 0 < t < T/2$. The output is written in a file 'cap3.out'

```
sum[0]=bin[0]*bin_width;
for(j=1;j<bin_nu;j++)
{
sum[j]=sum[j-1]+bin[j]*bin_width;
}
for(j=0;j<bin_nu;j++)
{
fprintf(fp3,"%+8.3f      %+8.3f\n",sum[j],
                          (j-(bin_nuf)/2+0.5)*bin_width);
}
for(j=0;j<bin_nu;j++)
{
fprintf(fp3,"%+8.3f     %+8.3f\n",sum[j]+sum[bin_nu-1],
                           -(j-(bin_nuf)/2+0.5)*bin_width);
}
for(j=0;j<bin_nu;j++)
{
fprintf(fp3,"%+8.3f     %+8.3f\n",sum[j]+2*sum[bin_nu-1],
                           (j-(bin_nuf)/2+0.5)*bin_width);
}
for(j=0;j<bin_nu;j++)
{
fprintf(fp3,"%+8.3f     %+8.3f\n",sum[j]+3*sum[bin_nu-1],
                          -(j-(bin_nuf)/2+0.5)*bin_width);
}
}
```

[*] Electronic address: xarvind@andrew.cmu.edu; Present address: Department of Physics, Carnegie Mellon University, Pittsburgh PA 15217, USA.
[†] Electronic address: chandi˙p@rediffmail.com
[‡] Electronic address: ravics@yahoo.com
[§] Electronic address: indu@imsc.res.in
[¶] Electronic address: shankar@imsc.res.in
[1] B. L. Saraf, *Physics Through Experiments Vol. I & II*, Vikas Publishing House, New Delhi, 1979.
[2] K. K. Gan, H. P. Kagan, and R. D. Kass, Simple demonstration of the central limit theorem using mass measurements, American Journal of Physics **69**(9), 1014–1020 (September 2001).
[3] E. M. Levin, Experiments with loaded dice, American Journal of Physics **51**(2), 149–152 (February 1983).
[4] G. Fischer, Exercise in Probability of and statistics or probability of winning at tennis, American Journal of Physics **48**(1), 14–19 (Jan 1980).
[5] P. C. B. Fernando, Experiment in elementary statistics, American Journal of Physics **44**(5), 460–463 (May 1976).
[6] N. B. Tufillaro, Generating a fractal using a capacitor, American Journal of Physics **69**(6), 721–22 (June 2001).
[7] D. T. Gillespie, A Theorem for physicists in the theory of random variables, American Journal of Physics **51**(6), 520–532 (June 1983).
[8] J. D. Ramshaw, Probability densities and the random variable transformation theorem, American Journal of Physics **53**(2), 178–180 (February 1985).
[9] M. D. Sturge and S. B. Toh, An experiment to demonstrate the canonical distribution, American Journal of Physics **67**(12), 1129–31 (December 1999).
[10] R. Aguayo, G. Simms, and P. B. Siegel, Throwing nature's dice, American Journal of Physics **64**(6), 752–758 (June 1996).
[11] H. W. Lewis, What is an experiment?, American Journal of Physics **50**(12), 1164–1165 (December 1982).
[12] H. W. Lewis, What is an experiment? II, American Journal of Physics **53**(6), 592–593 (June 1985).
[13] C. S. Barnett, Probabilistic description of radioactivity based on the good-as-new postulate, American Journal of Physics **47**(2), 173–177 (February 1979).
[14] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipies C*, Cambridge University Press, 2 edition, 1992.