Tom, Dick, and Mary
Discover the DFT

The story of three students’ discovery of the relationship between continuous and discrete transforms

J. R. DELLER, JR.

Discrete Fourier techniques are increasingly being taught as material detached from fundamental continuous-time Fourier analysis. The student is left with an unclear understanding, if any, of the very significant relationships between magnitude and phase spectra generated digitally, and the continuous-time signal which is being analyzed. This article tells the story of three undergraduate students who discover the DFT, armed only with a knowledge of analog methods.

Smaller and Smaller Lies

A professor once said to me “Education is the process of telling smaller and smaller lies.” I wish I could remember who it was so I could give proper attribution for that wonderful quote which I have used so many times in my own teaching. Educators (whether they be professional or otherwise) do not ordinarily tell malvolent “lies.” More accurately put, the state of understanding and knowledge of the learner does not always permit the whole “truth” to be told. By simplifying and omitting details, the teacher lays a foundation for the next levels (and there may be several) of understanding, the more complex details of which can be presented when the student is ready to grasp them. The simplified, alleviated-detail, versions of the subject sometimes contain “lies” of omission or inaccuracy for the sake of simplicity. The process is recursive; at each level the “lies” get smaller.

At some point, all of us stop learning about any given subject. Inevitably, the learning stops before we achieve the consummate “truth” about most subjects. In engineering, this can have unfortunate consequences for design and innovation. It is particularly unfortunate,
However, when the incomplete version has unnecessarily supplanted the complete understanding; that is, when the student could have just as easily grasped the complete picture, but is left with an incomplete understanding.

Such is the case with the discrete-time Fourier transform (DTFT) and the discrete Fourier transform (DFT). These transforms are at the heart of modern spectral analysis, and since they are almost invariably used to analyze problems arising in continuous time, it is of paramount importance that the analyst recognize the relationships between discrete and continuous spectral techniques. While the historical development of discrete Fourier techniques is rich and detailed \([1, 2, 3]\) the fundamental roots of the methods lie in the Fourier series, as is clearly brought out in the often-cited paper by Cooley and Tukey \([4]\). In this sense, the essence of the DFT was discovered by Fourier in 1807 \([5, \text{p. 431}]\), long before the conception of the digital computer. Earlier in this century, Fourier’s ideas were brought to bear on the contemporary sampled-data problem by Shannon in the United States \([6, 7]\) and Kotel’nikov in the Russian literature \([6, 8]\). However, much of the original motivation for the development of discrete Fourier techniques has been lost in recent tutorial literature.

The January, 1992 issue of *IEEE Signal Processing Magazine* celebrates the 25th anniversary of the FFT. The year 1967 coincides roughly with the genesis of the modern discipline we know as “digital signal processing.” Somewhere along the way, we have lost the early understanding of the DFT. Respected teachers and textbook writers have found it expedient to implicitly treat the DTFT and DFT as entities apart from the more fundamental principles upon which they are built. At best, some obtuse attempts may be made to tie the subjects together using “impulse sampling,” a “lie” from which many students never recover. It is assumed that the student, at the next level of understanding, will make the fruitful ties back to the classical theory which enriches the utility and meaningfulness of the discrete transforms. Regrettably, this next level of understanding may not be realized for many practitioners, resulting in a career of spectral analysis based on an incomplete understanding. In my own teaching, I feel uncomfortable leaving the students with the “lie” that the DTFT and DFT are something different from what they already know about continuous transforms. Fortunately, the following series of events occurred in a class I was recently teaching. If you doubt the veracity of this story, please reread the opening sentence of this article.

### Too Much Homework

Tom, Dick, and Mary were juniors taking a course in signal and linear system analysis. They were just completing the part of the course covering Fourier analysis of continuous signals, and were working together on an assignment with a seemingly countless number of problems involving the plotting of magnitude and phase spectra based on the Fourier transform.

“Too much homework,” said Tom. “We’ll never get this all done by hand.”

Dick suggested using one of the personal computers in the next room to “at least do the plots.” Mary pointed out that the plots were continuous curves and that they could at best plot samples of the spectra. She suggested that if they could find closely spaced samples, then maybe they could “connect the dots” to get at least an approximation to the desired spectra. Soon they were staring at the first signal \(x_1(t)\) and its Fourier integral:

\[
X_1(f) = \int_{-\infty}^{\infty} x_1(t) e^{-j2\pi ft} \, dt
\]

wondering how they would be able to compute samples of \(X_1\) using discrete computations.

“We can’t do an integral on the computer even if we just want values of \(X_1(f)\) at samples of \(f\),” Tom correctly noted.

They thought for a while and almost abandoned the computer when Mary recalled a point which ultimately saved the evening. She remarked that the Fourier series (FS) resulted in a sort of sampled frequency domain. Dick thought that was irrelevant because “we are not working on FS problems—\(x_1(t)\) is not a periodic signal,” Dick observed, “so I don’t see how we can apply the FS.”

Mary pointed out that only one period of a periodic function is ever used in constructing a FS. That prompted Tom to suggest, “Let’s make \(x_1(t)\) periodic, and see what we can do with it.”

Dick was still skeptical, but he agreed to go along with the plan. “What period should we give it?” he asked. “I don’t know,” admitted Mary, “but let’s try a general period, say \(T_y\), and call the periodic signal \(y(t)\). Then we’ll have this,” Mary said, writing on the blackboard.

\[
y(t) = \sum_{k = -\infty}^{\infty} x_1(t - k T_y)
\]

She then sketched the signal \(x_1(t)\) (Fig. 1a), noting that it has a “time width” \(T_1 = 8\). She also sketched two cases of \(y(t)\): Case 1 in which \(T_y > T_1\) (Fig. 1b); and Case 2 in which \(T_y \leq T_1\) (Fig. 1c).

“I’m not sure where this is going,” said Tom, “but let’s write the FS for \(y(t)\).” He wrote

\[
y(t) = \sum_{m = -\infty}^{\infty} c_m e^{j2\pi m f_y t}, \quad f_y = \frac{1}{T_y}
\]

By using this definition of the Fourier transform, rather than the similar one based on radian frequency, \(X_1(\omega) = \int e^{j\omega t} \, dt\), we avoid some scale factors in the following discussion. The developments remain essentially unchanged if the radian-frequency definition is used.
Then he realized they were onto something. "These coefficients are just the samples of the FT of the original signal (with a scale factor)," he said as he wrote \( \frac{1}{T_y} X_1(m f_s) \) at the end of the last line.

"It looks like we've gotten the samples in the frequency domain that we want," observed Mary. "Let's summarize. If we take samples of the Fourier transform \( X_1(m f_s) \) with \( f_s = \frac{1}{T_y}, T_y > T_1 \), then scale them by \( f_s = \frac{1}{T_y} \), we have the FS coefficients for a periodic version of \( x_1(t) \) whose "copies" do not overlap. We have the samples of the FT that we want, and apparently haven't lost any information since we can recover \( X_1(t) \) from the samples by constructing the periodic waveform using Eq. 3, then taking the 'primary' period."

"What would happen if we tried the same trick with frequency samples taken at \( f_s = \frac{1}{T_y} \) with \( T_y \leq T_1 \)?" Tom wondered. "I'll bet those samples are still FS coefficients, but for the 'overlapped' version of \( y(t) \) in Case 2." After playing with the problem for a while, the group confirmed that this was the case. They knew that a periodic function did not have a valid FT, but that one could be constructed using impulse functions. They had been warned about the mathematical hazards of impulse functions, but, for lack of anything better to do, they finally resorted to taking the FT of the signal they had constructed using the FS. Since they weren't sure that this manufactured signal were truly \( y(t) \), they called it \( y'(t) \), and wrote
\[ y'(t) = \sum_{m=-\infty}^{\infty} \frac{1}{T_y} X_1(m f_s) e^{j2\pi m f_s t} \]

to obtain
\[ y'(f) = \frac{1}{T_y} \int_{-\infty}^{\infty} X_1(m f_s) e^{j2\pi m f_s f} e^{-j2\pi f t} \, dt \]
\[ = X_1(f) \sum_{m=-\infty}^{\infty} \delta(f - m f_s) \]

Now using the fact that \( \sum_{m=-\infty}^{\infty} \delta(f - m f_s) \) is the FT of the time signal \( T_y \sum_{m=-\infty}^{\infty} \delta(t - m T_y) \) [9, p. 723], and the property that multiplying two FTs corresponds to convolution of their time functions, they wrote,
\[ y'(t) = \int x_1(t - \tau) \sum_{m=-\infty}^{\infty} \delta(\tau - m T_y) \, d\tau \]
\[ = \sum_{m=-\infty}^{\infty} x_1(t - m T_y) = y(t) \]

"OK!" exclaimed Mary, "Now we know that we can always obtain at least a periodic version of \( x_1(t) \) using scaled samples of \( X_1(f) \) as FS coefficients. If we want one period of \( y(t) \) to be exactly \( x_1(t) \), we must make sure that the frequencies of the samples of \( X_1(f) \) at \( m f_s = m T_y \) are close enough together, \( T_y > T_s \)."

Dick remarked that they were still a long way from being able to plot spectra on the computer. "To work with these spectra on the computer, we must get rid of the continuous signals in both time and frequency domains," he said. "At least we know what we can do with frequency samples," he conceded.

"We talked a lot about time-frequency duality in class," recalled Tom, "and it seems that we should be able to 'reverse' the process we've just used in order to get samples in the time domain."

"Yeah," agreed Mary, "I'll bet there's a general principle here: If we're willing to make one domain periodic, all the information will be contained in samples of the other. But, the
samples in one domain have to be close enough to keep the periods from overlapping in the other (if that’s possible). Let’s try Tom’s idea.”

Since they didn’t know what $X_1(f)$ looked like, Mary sketched a hypothetical version of $X_1(f)$ on the board (Fig. 2), carefully labeling its one-sided bandwidth as $\omega_s$. She then drew two periodic functions, both with period $f_s$, and both labeled $Y(f)$. In Case 1, the period $f_s > 2f_s$, and in Case 2, $f_s = 2f_s$.

The group paused for a moment to consider an important detail. Tom objected to the sketches on the grounds that $X_1(f)$ is a complex function which could not be drawn in two dimensions. However, Mary correctly pointed out that the signal $x_1(t)$ is real and even in this case, implying a real and even FT. “We’re just lucky here,” Mary noted, “but I guess in general we have to be careful with such pictures because we are really adding complex numbers.”

“I guess we could still get the idea of frequency overlap using these diagrams,” Dick observed. Dick was correct, but the magnitude spectra do not simply add directly as we might infer from these pictures, I told them the next day as we discussed this point in my office. Textbooks do not always make this point clear, and we must be careful with such sketches.

The other detail that they neglected is the fact that a signal cannot be both time-limited, as in the case with $x_1(t)$, and frequency bandlimited, as implied by Mary’s sketch of $X_1(f)$ [9, p. 154]. This negligence turned out to be beneficial, for, had they remembered this fact, they might have been discouraged from making $X_1(f)$ periodic. We discussed this point at some length the following day.

The students began to scrutinize the cases of Figs. 2b and 2c. In either case, they speculated that the periodic function could be represented by a “FS” whose coefficients were related to samples of time. It was simply a matter of reversing the roles of time and frequency in the analysis. Dick guessed correctly that the time samples would be $x_1(nT_s)$, with $T_s = 1/f_s$, by making an analogy to what happened in the dual problem.

Following the analysis in the frequency-sampling problem, the group quickly discovered that their intuition was correct. The periodic (and complex) frequency function $Y(f)$ could be represented by a “FS.” They first computed the “FT” for the “signal” $X_1(f)$, calling it $X_1'(t)$,

$$X_1'(t) = \int X_1(f) e^{-j2\pi ft} df$$

They noted that this “FT” is very close to the time signal,

$$x_1(t) = \int X_1(f) e^{j2\pi ft} df = x_1'(-t)$$

but agreed to ignore this fact temporarily in order to preserve the analogy to the frequency-sampling case. They carefully wrote down the “FS” for the periodic function (Eq. 10), letting $f_s$ play the role of “time,” and $\omega_s$ play the role of the fundamental “period.” Accordingly, $1/\omega_s = T_s$ is the analogous quantity to fundamental “frequency.” In these terms, the “FS” is

$$Y(f) = \sum_{n=-\infty}^{\infty} \beta_n e^{j2\pi n\omega_s t}, \quad T_s = 1/\omega_s$$

$$\beta_n = \frac{1}{f_s} \int_{-f_s/2}^{f_s/2} Y(f) e^{-j2\pi n\omega_s t} df = \frac{1}{f_s} X_1'(nT_s) = \frac{1}{f_s} x_1'(-nT_s)$$

The “FS coefficients are scaled samples of the (reversed time) waveform $x_1'(t) = x_1(-t)$ so

$$Y(f) = \sum_{n=-\infty}^{\infty} \frac{1}{f_s} x_1'(-nT_s) e^{j2\pi n\omega_s t}$$

The group decided they didn’t like using the samples in reversed time, so they made the simple change of replacing $n$ by $n$, to obtain
"We've got it!" Mary said excitedly. "As long as we take \( f_s = 1/T_s > 2f_b \), then the first period of \( Y(f) \) will be exactly \( X(f) \)."

"But we need samples of \( X(f) \)," Dick reminded her, "in order to plot them and work with them on the computer."

"No problem," said Tom, "we simply calculate \( Y(f) = X(f) \) at any frequency we want. For example, take the frequency \( f_0 \)," and he wrote on the board

\[
Y(f_0) = X(f_0) = \frac{1}{f_s} \sum_{n=-\infty}^{\infty} X(nT_s) e^{-j2\pi n f_0 T_s}
\]

"This is valid for any frequency sample \(-f_b \leq f_0 < f_b \) since that's the range over which \( Y(f) = X(f) \)."

Using sampled values of the given signals, and the expression they had derived in Eq. 17, Tom, Dick, and Mary were able to computer generate the plots for each of their homework problems. Not knowing the bandwidth, \( f_b \), for any of the signals, they had to guess the sample period \( T_s \) to use in each case. They arbitrarily chose \( T_s = 1/f_s = 1 \) s for all signals. In some cases, this was satisfactory; in others, quite unsatisfactory. The magnitude spectrum for result for signal \( x_i(t) \) is shown in Fig. 3. (Note that the phase is either zero or \( \pi \) radians at every sample, since the signal \( x_i(t) \) is real and even, resulting in a real, even FT.) Figure 3 was obtained using Eq. 17 at the frequency samples \( kf/32 = k32 \) for \( k = 0, 1, \ldots, 16 \).

When we met in my office the next day, Tom, Dick, and Mary were justifiedly excited about their discovery.

Gold Stars: Part I

I commended them for their innovation, and took the opportunity to explain just how profound their work was. "First of all," I told them, "in the time sampling case, what you have discovered is the essence of a fundamental result in information theory known as Shannon's Sampling Theorem [6,7]. This principle says that if we sample a time waveform at a high-enough rate, \( f_s = 1/T_s > 2f_b \), then a complete continuous-time signal like \( x_i(t) \) is recoverable from the time samples \( x_i(nT_s) \).

In theory, the method for doing so is obvious from your work: Form the "FS" for \( Y(f) \) as in Eq. 16. The first period of \( Y(f) \) is the FT \( X_i(f) \), so we can inverse FT this initial period to recover the signal \( x_i(t) \). In practice, we apply a lowpass filter to recover the signal occupying the primary frequency band.

"This lowpass filtering operation can be shown to be a discrete convolution operation involving the samples of the signal as inputs to the filter. The process is usually called interpolation, since it allows us to interpolate between sample values in the signal."

"As long as you sample \( x_i(t) \) fast enough, the FS \( Y(f) \) is an excellent candidate for spectral analysis of the original signal, since it represents an exactly periodic version of \( X_i(f) \). But even when samples are not taken fast enough (or cannot be because of infinite bandwidth), \( Y(f) \) is closely related to \( X_i(f) \) in that it is a periodic, but aliased, version of \( X_i(f) \). The term aliasing refers to the overlap in the frequency spectrum which causes frequencies above \( f_b \) in the signal to be confused with those in the band \( 0 \leq f < f_b \)."
We studied two signals in their homework in some detail to understand this point. First, I generated a plot representing a very good approximation to the true continuous-time magnitude spectrum of signal \( x_1(t) \) (Fig. 4a) for the sake of discussion. "You were lucky in this case," I told them, "since there is energy in this signal beyond your assumed bandwidth of \( f_s/2 = 1/2 \). This is more evident on the log scale (Fig. 4b). However, the amount of energy is negligible compared to that in the range \( 0 \leq f < 1/2 \), so there is little distortion of the spectrum on the range you plotted. (Imagine what would happen if Fig. 4a or Fig. 4b were made periodic with period \( f_s = 1 \) as above.) However, there is some aliasing as we see, for example, by comparing your sample at the frequency \( f = 0.3125 \), with the same frequency on the "true" spectrum. We see that, on your plot, \( |X_1(0.3125)| = 0.723 \), whereas on the "true" spectrum, \( |X_1(0.3125)| = 0.520."

So we concluded that we will not always be able to "guess" a reasonable sample rate. The most flagrant example of aliasing occurred in the next problem which involved the signal

\[
x_2(t) = e^{-0.03t} \sin(5t) [u(t - 1) - u(t - 93.1)]
\]

(18)

with \( u(t) \), the unit step function, defined to be unity for \( t \geq 0 \) and zero otherwise. Based on our previous discussion, the group now understood that a sample rate of \( f_s = 1 \) was inappropriate since the energy in this damped sinusoid is concentrated around the frequency \( f_0 = 5/2 \pi = 0.796 \text{Hz}. "A sampling rate of } f_s = 2 (0.796) = 1.6 \text{ Hz is needed to prevent 'overlap' in the periodic FT," Tom noted correctly.

"What happened in this case is very revealing," I told them. "Look at your spectral magnitude plot (Fig. 5a). It seems to indicate energy at \( f = 0.2 \text{ Hz}, but fails to show any energy at the frequency of the sinusoid (which is not even in the range of your plot). What happened?"

I suggested that they extend their plot to include one or more periods of the "overlapped" FT. Using the program
they had written the night before, they did so (Fig. 5b),
knowing that the result would be periodic with period \( T_1 = 1 \). 
While they were doing that, I plotted the true spectrum of 
\( x_3(t) \) (Fig. 5c).

"I see what happened," said Tom. "The bogus peak we got 
in the range \( 0 \leq f \leq 1/2 \) is the result of one of the copies of the 
true spectrum that got shifted into that range." They identified 
the "correct" peak in the periodic spectrum and convinced 
themselves that this problem would not have occurred if a 
proper sampling rate had been used.

"Now let me tell you a further significance of your result," 
I continued. "The 'FS' \( Y(f) \) you've discovered is very close to 
what is known as the discrete-time Fourier transform (DTFT) 
for the discrete time signal consisting of the samples 
\( x_3(nT_s) \), \( n = -1, 0, 1, 2, \ldots \). Usually, books will tell you that 
the DTFT is the something like the following." I wrote on the board

\[
Y_{DTFT}(f) = \sum_{n=-\infty}^{\infty} x_3(nT_s) e^{-j2\pi fnT_s}
\]  

(19)

"That's just our 'FS' except for a scale factor \( 1/T_1 \)," Mary 
said. "Yes," I agreed, "which happened to be unity in your 
numerical work anyway. So can you tell me the significance of 
the DTFT? Think about what we've just worked through."

"Well, we wanted to be able to plot spectra using the 
computer, so we had to have discrete samples in both 
domains," Tom said. "Yes," I interjected, "and more generally, 
computers are being used to replace all kinds of analog 
systems and filters so that the computations, and in this case, 
spectral analysis, must be done with discrete samples. Usually, 
however, the samples relate to continuous phenomena 
since most signals are continua. So we'd like to be able to 
perform "meaningful" spectral analysis using the computer."

"You mean we want to be able to relate the results to the 
continuous world, since that's where they really come from," 
Mary remarked.

"Precisely."

"So the DTFT is very important because it allows us to 
compute a periodic replica of the continuous FT using just 
time samples," Dick noted.

"As long as we sample fast enough in time," I added. 
"Now what is the effect of the scale factor omission in the 
definition of the DTFT?" I asked.

"No big deal," said Tom, "it just means that the DTFT is 
a scaled periodic replica of the original FT." Looking back at 
his notes from the night before (Eq. 10), he wrote

\[
Y_{DTFT}(f) = f_s \sum_{n=-\infty}^{\infty} x_3(f-nf_s) = \frac{1}{T_s} \sum_{n=-\infty}^{\infty} X_1(f-nf_s)
\]  

(20)

"So is the DTFT something new?" I asked them. "Or is it 
a natural extension of classical theory?" The question was 
rhetorical, and they were very proud of their accomplishment.

"When you see the DTFT in your work with discrete-time 
signals, you will forever remember its origins since you 
worked so hard to discover them."

"Now I have a challenge for you," I said after they had 
congratulated themselves. Tom, Dick, and Mary looked 
astonished since they had been awake most of the night facing 
what they considered to be a more than sufficient challenge.

"In the first part of your work, you show how to use a set of 
frequency samples to represent a continuous-time waveform. 
In the second part, you show how to use time samples to 
represent a continuous-frequency function. (Then you use 
that result to compute frequency samples. To use a computer 
for spectral analysis, it would be useful to have a transform 
which goes back and forth between sets of samples. You are 
close. See if you can find such a transform.)"

"Let me give you a hint," I added. "You notice that your 
DTFT computation uses a doubly-infinite number of time 
samples."

"Yeah," Tom agreed, "we had to chop off the signal after 
a large number of samples if the original signal was not 
'time-limited'. I mean if \( T_1 = \infty \). So I guess we only got an 
approximate spectrum."
"Yes, and that will always be the case in discrete-time analysis," I pointed out. "You can never enter an infinite number of time samples. There are smart ways to truncate long signals to get better estimates of the spectra, but that's another story [10, 11]. Let's just assume that we have a signal which starts at time \( t = 0 \) and 'ends' at some finite time. An example is \( x_3(t) \) in your homework, but the next one, \( x_4(t) \), is simpler:

\[
x_3(t) = (0.9)^{1000t} (u(t) - u(t - 0.0151))
\]

so \( x_3(t) \) is nonzero on \( 0 \leq t < 15.1 \text{ ms} \).

Suppose \( T_s \) is the sampling period you choose and that you get \( N \) samples on the nonzero range, \( x_3(0), x_3(T_s), x_3(2T_s), ..., x_3(N - 1)T_s \). These samples comprise the nonzero part of the discrete-time signal \( x_3(nT_s) \)." I sketched the samples on board for \( T_s = 1 \text{ ms} \) (Fig. 6).

"Is \( x_3(t) \) bandlimited to some finite \( f_s \)?" they wanted to know. "And, if so, is \( f_s > 2/f_s \)?"

"\( x_3(t) \) can't be bandlimited because it is of finite time duration. However, it can be assumed to be practically bandlimited to some finite \( f_s \), either inherently or because of a pre-processing lowpass filtering operation. For your theoretical work, assume that there is practically \( f_s < \infty \) but that you don't know it's value," I responded. "This means that you have to be prepared for some aliasing since you don't know the proper value of \( f_s \)."

"I appreciate what you did last night! Now see if you can take it to the next logical step."

I'd like to tell you that Tom, Dick, and Mary excitedly left my office, eager to tackle this new challenge. However, this is a true story. Something unfit for print was muttered in the hallway as they walked away.

### Onward to the DFT

The three students puzzled over Eqns. 7 and 9 for a while thinking maybe they could just use a similar equation to Eq. 7 at times \( t = nT_s \) to get the time samples they wanted. "But we don't really have \( X_3(f) \) to sample," Tom pointed out. "What we have is a periodic version of \( X_3(f) \) that we again call \( Y(f) \). In this case, since we have only \( N \) time samples, Eq. 16 can be simplified." He wrote

\[
Y(f) = \frac{1}{f_s} \sum_{n=0}^{N-1} x_3(nT_s) e^{-j2\pi nT_s f}
\]

"Of course, we can compute any samples of \( f \) that we want, but which ones?" Dick wondered. They pondered this for a long time. Mary reiterated the point that the signal \( x_3(t) \) could not be frequency bandlimited because it is time-limited. Nevertheless, she suggested that they go back and look at what happened in their..."
first analysis in which \( X_i(f) \) was assumed (incorrectly) to be frequency-limited to \( f_b \). “Look at Eqs. 7 and 9,” she instructed her classmates. “If we had samples of \( X_i(t) \) to work with, and if we took these samples at spacing \( f_s = 1/NT_s = f_b/N \), then Eq. 7 would produce, according to Eq. 3:”

\[
y(t) = \sum_{k=-\infty}^{\infty} x_3(t - kNT_s)
\]

(23)

“And, if we used Eq. 7 at times \( nT_s \), then we would get the numbers:”

\[
y(nT_s) = \sum_{k=-\infty}^{\infty} x_3(nT_s - kNT_s)
\]

(24)

They sketched the sequence \( y(nT_s) \) on the blackboard (Fig. 7a) and discovered that the first \( N \) points were exactly the sequence \( x_3(nT) \). In fact, they varied the value of \( f_s = 1/T_s \), and discovered that using a smaller value of \( f_s = 1/NT_s \) would cause “extra space” (zeros) between the copies of \( x_3(t) \) in \( y(t) \) (Fig. 7b), and “overlap” (time aliasing) of the copies would occur if \( f_s \) were chosen too large, \( f_s > 1/T_s \) (Fig. 7c).

“But we don’t have the samples of \( X_3(t) \) to work with,” noted Dick. “We have only a periodic, maybe aliased, version of \( X_3(t) \).”

“I have a hunch that samples of \( Y(f) \) taken at the right spacing \( f_s \leq f \leq f_s \) will still give us back the periodic sequence (Eq. 24),” Tom said. “Something like what happened when we sampled the aliased periodic FT last night (Eqns. 7-9),” he added with some uncertainty in his voice.

They knew that this case had to have something to do with the first analysis they had done the night before concerning frequency sampling. They began to pore over those results. “Look at the pattern we followed there,” said Mary. “We had a bandlimited signal \( x_3(t) \) and we let it become periodic so that we could discretize the frequency domain. I think we want to do the same thing here. Forget the fact that \( Y(f) \) is periodic. That’s what’s throwing us off. Just look at one period of it,” and she wrote on the blackboard the following definition:

\[
Y_3(f) = \begin{cases} 
Y(f), & \frac{f_s}{2} \leq f \leq \frac{f_s}{2} \\
0, & \text{otherwise}
\end{cases}
\]

(25)

and drew the sketch in Fig. 8. “Of course, we’ll let \( y_3(t) \) be the corresponding time signal. I’m not sure what \( y_3(t) \) is, but I know we will make it periodic if we build a FS out of samples of \( Y_3(f) \), and, if Tom’s hunch is right, the signal \( y_3(t) \) will match \( y(t) \) at the sample times \( nT_s \),” Mary deduced.

“By chopping off \( Y(f) \) to create \( Y_3(f) \),” Dick pointed out, “we force \( y_3(t) \) to be a continuous-time signal.” He also observed that \( y_3(t) \) would be exactly \( x_3(t) \), and \( Y_3(f) = X_3(f) \) if the original signal, \( x_3(t) \) had been bandlimited and sampled fast enough.

“Right on both counts,” said Tom, “but Mary is considering the general case in which \( Y(f) \) and \( Y_3(f) \) might involve some aliasing. I also see what she’s trying to do by making \( y_3(t) \) a continuous-time signal. She’s taking us back to the very first problem we studied last night—continuous-time, discrete-frequency.”

“I see where we’re going,” responded Dick, “but if we follow the approach we did last night, we’ll end up writing a FS for a periodic version of \( y_3(t) \), not \( y(t) \). Then we’ll evaluate at discrete times, and end up with samples of a periodic version of \( y_3(t) \), not samples of \( y(t) \), and certainly not the samples \( x_3(nT) \).”

“Go ahead and do it anyway,” Tom persisted. “I think it will work.” So they began to meticulously follow the steps that had become familiar by now. First they created the periodic extension of the somewhat mysterious signal \( y_3(t) \):

\[
w(t) = \sum_{i=-\infty}^{\infty} y_3(t - iT_w)
\]

(26)

anticipating that they would sample \( Y_3(f) \) at spacing \( f_w = 1/T_w \). They all agreed that \( T_w \) would have to be some multiple of \( T_s \) because the period would ultimately have to be some integer number of samples, say,

\[
T_w = MT_s \text{ or } f_w = \frac{1}{MT_s}
\]

(27)

Then they wrote the FS:

\[
w(t) = \sum_{k=0}^{\infty} y_k e^{j2\pi ft \cdot}, \quad f_w = \frac{1}{T_w}
\]

(28)

\[
y_k = \frac{1}{T_w} \int_0^{T_w} w(t) e^{-j2\pi ft} \, dt
\]

(29)

They knew from past experience that whether \( w(t) \) were
an "overlapped" version of $y_3(t)$ or not (see Eqs. 5-9),

$$\gamma_k = \frac{1}{T_w} Y_3(kf_w)$$

Therefore,

$$w(t) = \frac{1}{T_w} \sum_{k=-\infty}^{\infty} Y_3(kf_w) e^{2\pi j k f_w t} \tag{30}$$

They noted that they could replace $Y_3(f)$ by $Y(f)$ by restricting the summation:

$$w(t) = \frac{1}{T_w} \sum_{k} Y(kf_w) e^{2\pi j k f_w t} \tag{32}$$

They now had frequency samples on the right side with which they could compute time samples. In particular, they wanted samples at times $nT_s, \quad n = 0, 1, \ldots, N-1$. So they evaluated $w(t)$ at the discrete times,

$$w(nT_s) = \frac{1}{T_w} \sum_{k=-\infty}^{\infty} Y(kf_w) e^{2\pi j k f_w nT_s} \tag{33}$$

"Great," muttered Dick, sarcastically, "we can compute samples of some signal we don't understand or want."

"Hold on," said Mary. "These are samples of a periodic version of $y_3(t)$. We just have to figure out how $y_3(t)$ is related to our original time samples."

"Look," she said, as she wrote

$$y_3(nT_s) = \int_{-f_s/2}^{f_s/2} Y_3(f) e^{2\pi j f nT_s} df \tag{34}$$

"Hey, that just looks like a FS coefficient computation for the periodic $Y(f)$," exclaimed Tom.

Mary was busy rifling back through her notes from the night before. "Yeah, we've seen that before!"

After a brief search she rediscovered Eq. 14 and wrote, triumphantly,

$$y_3(nT_s) = f_s \beta_{-n} = x_3(nT_s) \tag{35}$$

then,

$$w(nT_s) = \sum_{i=-\infty}^{\infty} x_3(nT_s - iT_w) \tag{36}$$

Recalling their agreement (Eq. 27), she finally wrote

9. Samples of the (a) magnitude, and (b) phase, spectra of the discrete-time signal $x_3(nT_s)$ of Fig. 6 obtained using Eq. 39. (c) Magnitude spectrum of the continuous-time signal $x_3(t)$ of Eq. 21.
\[ w(nT_s) = \sum_{i=-\infty}^{\infty} x_3(nT_s - iMT_s) \]

"Voila! \( w(nT_s) \) is just a periodic, possibly aliased version of \( x_3(nT_s) \)," Tom said excitedly. "As long as we take \( M \geq N \), then the samples \( w(0), w(T_s), \ldots, w((N-1)T_s) \) will be the time samples we are looking for!" Referring back to Eqs. 32 and 27 with \( M = N \), he wrote

\[
x_3(nT_s) = \frac{1}{N T_s} \sum_{-\frac{f_s}{2} \leq \frac{k f_s}{N} \leq \frac{f_s}{2}} Y \left( \frac{k f_s}{N} \right) e^{j2\pi \frac{k f_s}{N} nT_s} \]

\[ n = 0, 1, \ldots, N-1 \]

"That gives us the \( N \) time samples in terms of \( N \) frequency samples," said Dick. "Now we need to get the \( N \) frequency samples using the \( N \) time samples. I think we can use the result we used for the plots in our homework."

They dug back through their notes and found Eq. 16. Noting that \( x_3(nT_s) \) is zero except for the first \( N \) points, they wrote

\[
Y \left( \frac{k f_s}{N} \right) = \frac{f_s}{N} \sum_{n=0}^{N-1} x_3(nT_s) e^{-j2\pi n k f_s / N}, \quad \frac{f_s}{2} \leq \frac{k f_s}{N} \leq \frac{f_s}{2} 
\]

They computed the 15 samples in frequency using Eq. 39 and the corresponding magnitudes and phases shown in Figs. 9a and 9b, respectively. By hand, they plotted the true magnitude spectrum (shown in part (c) of the same figure) to see how closely their samples matched the true spectrum.

Gold Stars: Part II

With understandable pride, the students brought me Eqs. 38 and 39 and their plots the next day. I applauded their efforts and suggested we make a few minor adjustments.

"Your Eq. 39 is used to compute samples of \( Y(f) \). Suppose we wanted to compute samples of the more conventionally used ‘DTFT’ that we discussed yesterday. What would that look like?"

They reviewed our discussion of the DTFT and realized that only a scale factor change was necessary. Since \( Y_{DTFT}(f) = f_s Y(f) \), we wrote

\[
Y_{DTFT} \left( \frac{k f_s}{N} \right) = \sum_{n=0}^{N-1} x_3(nT_s) e^{-j2\pi n k f_s / N} 
\]

The inverse relation also had to account for the scale factor,

\[
x_3(nT_s) = \frac{1}{N} \sum_{n=0}^{N-1} Y_{DTFT} \left( \frac{k f_s}{N} \right) e^{j2\pi k N n} \]

"You have also done the logical thing of sampling the FT (or now DTFT) over the ‘primary band’. That is, you took the appropriate samples on the frequency range \( \frac{f_s}{2} \leq f \leq \frac{f_s}{2} \). The problem with this is that the values of \( k \) (when written in general terms) depend on whether \( N \) is odd or even. When \( N \)
is odd, then the values are \( k = -\frac{N-1}{2}, \ldots, 0, \ldots, \frac{N-1}{2} \); but when

\( N \) is even, \( k = -\frac{N}{2}, \ldots, 0, \ldots, \frac{N}{2} - 1 \).

"That means we have to have two different formulas for the frequency samples," observed Tom.

"Right. Unless we can find a way around the problem. What signal processing people usually do is exploit the fact that the numbers \( Y_{DTFT} \) are periodic with period \( N \). (Remember that the DTFT has period \( f_n \), which corresponds to \( N \) samples.) Let's show this:

\[
Y_{DTFT}\left(\frac{(k+N)f_n}{N}\right) = Y_{DTFT}\left(\frac{kf_n}{N} + f_n\right) \quad (42)
\]

and

\[
e^{2\pi i (k+N)N} = e^{2\pi i kN} e^{2\pi i} = e^{2\pi i kN} \cdot 1 \quad (43)
\]

What this means is that, in either case, instead of computing \( Y_{DTFT}\left(\frac{kf_n}{N}\right) \) for negative values of \( k \), we can compute over nonnegative \( k \) values only. In either case, \( N \) odd or even, the values of \( k \) used are \( k = 0, 1, 2, \ldots, N-1 \). The students played with a few values of \( N \) to convince themselves that this was true. "Therefore, we can use Eq. 40 for these nonnegative \( k \)'s and modify Eq. 41 to read:

\[
x_3(nT_s) = \frac{1}{N} \sum_{k=0}^{N-1} Y_{DTFT}\left(\frac{kf_n}{N}\right) e^{2\pi i kN} n = 0, 1, \ldots, N-1 \quad (44)
\]

"The pair of relations (Eq. 40, used for \( k = 0, 1, 2, \ldots, N-1 \); and Eq. 44), are the essence of the discrete Fourier transform (DFT)," I elaborated. There is a group of efficient algorithms for computing the DFT relations that take advantage of the symmetry properties of the quantities involved. These algorithms are collectively known as the fast Fourier transform (FFT) (e.g. see [10-13]). We took a moment to compute the DFT of Eq. 40 and to plot the resulting magnitude spectrum (Fig. 10) and to verify that the inverse, Eq. 44, indeed produced the 16 samples of \( x_3(nT_s) \) shown in Fig. 6. The picture made clear that the DFT are samples of the first period of the DTFT on nonnegative frequencies, rather than the "primary period" that straddles \( f = 0 \).

"Let's review the significance of your DFT relations," I encouraged them. "We will always have to work with a finite-duration signal on the computer, like \( x_3(t) \), Dick began. "We then take \( N \) samples. The forward transform (Eq. 40) allows us to compute samples of the DFT." "What is the significance of those samples?" I asked. "Well, the DTFT will have to be an aliased version of the continuous-time FT, \( X_3(f) \)," Mary said. "\( x_3(t) \) cannot be bandlimited because it is time-limited."

"Right," I agreed. "Do you suppose this means that we can never use the DFT to study the continuous spectrum?"

"No, I guess you can always sample fast enough to make the aliasing insignificant," said Mary.

"Yes, or more likely, you will filter out all energy above \( f_s/2 \) and just resign yourself to studying the 'lower' frequencies," I added. "OK, so let's assume that one period of the DTFT is a very good approximation to \( X_3(f) \).

"Then the DTFT frequency samples will let us study the FT of the continuous-time signal using just additions and multiplications, no integrals!" said Tom. "And we can go back and forth between the frequency and time samples with the relations we derived."

"Yes, and you did that all with FS theory, which is in the spirit of the original developments of the transform," I reminded them. "So when you see the DFT in your later work, you will know that it is not a 'separate' transform, but that it has real meaning for analyzing continuous-time signals, which, after all, is what we are usually trying to do."

"Let's look at another important point which is inherent in your work," I encouraged them. "Assume that there is negligible aliasing so that Fig. 9c on \( 0 \leq f \leq f_s/2 \) represents the first half period of the DTFT except for a scale factor \( f_s \). Your samples on this range \( k = 0, 1, \ldots, 8 \) in either Fig. 9a or 10 compare favorably with the continuous spectrum. That is, they appear to be proper samples of the DTFT spectrum."

"But we don't see much of the detail in the spectrum," Tom pointed out. "Yes, that's exactly what I wanted you to notice," I said. "Is there a remedy for this problem? Remember how you sampled the FT at arbitrary frequencies in doing your homework plots (Eq. 17)?"

"I think the same basic idea applies here," said Mary. "If we rewrite the DFT relation (Eq. 40) as

\[
y_{DTFT}\left(\frac{N-1}{f_s/M}\right) = \sum_{n=0}^{N-1} e^{-2\pi i nT_s k f_s/M} x_3(nT_s) \quad (45)
\]

it is obvious that the \( k \)th result is the sample of the DTFT at frequency \( f = k f_s/M \). I guess we could compute these samples at whatever frequencies we want."

"That's right," I responded, "and a common thing to do is to compute these samples at the frequencies \( k f_s/M \) for \( k = 0, 1, \ldots, M-1 \) with \( M > N \). This gives us \( M \) equally (and more closely) spaced samples of the DTFT. \( M \) is frequently taken to be a power of two, since efficient and widely-available DFT (FFT) algorithms exist for this case. In this case, the DFT relation becomes

\[
y_{DTFT}\left(\frac{k f_s}{M}\right) = \sum_{n=0}^{M-1} x_3(nT_s) e^{-2\pi i nT_s k f_s/M} \quad (46)
\]

The last equality follows because \( x_3(nT_s) \) is assumed to be
zero outside the initial $N$ points. This is equivalent to taking the $M$-point DFT of the $N$-point signal $x_3(nT_s)$ to which $M - N$ zeros have been appended. You will get better resolution of the DFT corresponding to the assumed $N$-length signal $x_3(nT_s)$ using this process. If $x_3(nT_s)$ is, in fact, longer than $N$ points in duration, the only way to get both better resolution and a better DTFT to sample is to add more real data (not just appended zeros)."

We proceeded to plot the 64-point DFT of $x_3(nT_s)$ by appending 48 zeros to the original 16 points. The magnitude spectrum result is shown in Fig. 11. Better resolution of the continuous-time spectrum is evident in the result.

"Does Eq. 44 still return the correct signal $x_3(nT_s)$ when $M$ is used in place of $N$?" Dick wanted to know. "Good question," I responded. "The answer is yes, and it is not difficult to demonstrate by putting Eq. 46 into Eq. 44. So, again, we have a complete DFT pair with the ability to sample the DTFT for a finite-duration signal as closely as desirable."

"Finally, there's one more nuance that you need to be aware of," I cautioned the students. "It is customary in digital signal processing to 'normalize' the problem so that the sampling frequency is effectively unity, $f_s = 1$. Another way to look at this is that the time samples are indexed by integers, for example, $x_3(0), x_3(1), ..., x_3(N)$ so that the sample period, $T_s$, is effectively normalized to one. This reindexing has the accompanying effect of normalizing the frequency axis so that samples are taken in the interval $0 \leq f < 1$. Please notice that this process is not equivalent to sampling the original signal at a rate $f_s = 1$. Rather it amounts to reindexing samples that have been taken at an appropriate rate. You will explore these details more formally in later work in digital signal processing. However, I wanted you to be aware of this fact because you will often see the DFT relations written simply as, for example,

\begin{equation}
X_3(k) = \sum_{n=0}^{N-1} x_3(n) e^{-j2\pi nk/N}, \quad k = 0, 1, 2, ..., N-1
\end{equation}

and

\begin{equation}
X_3(n) = \frac{1}{N} \sum_{k=0}^{N-1} X_3(k) e^{j2\pi kn/N} \quad n = 0, 1, ..., N-1
\end{equation}

where $X_3(k)$ stands for what we have called $Y_{DTFT} \left( \frac{k f_s}{N} \right)$, or, since in this case $f_s = 1$, $Y_{DTFT} \left( \frac{k}{N} \right)$. Please note again that this is not the same thing as simply taking samples of the original time waveform at frequency $f_s = 1$.

The students were pleased with their discovery, and mentally and physically exhausted from their labors. They were packing up their books when I concluded the session. "I'm very proud of you for the hard work you did," I told them, "and since you seemed to enjoy this challenge so much, ..." But before I could finish the sentence, Tom, Dick, and Mary made a mad dash for the office door and were gone.

No More Lies

It is becoming increasingly likely that Tom, Dick, and Mary will spend an engineering career designing and analyzing spectra using only discrete techniques. The days of "pencil and paper" integral transforms as a basic design tools are receding into history. However, the students' basic notions of spectra, frequency, phase, energy, resonance, and so on, are all based in continuous-time concepts, and the problems they will work on are inevitably drawn from the continuous world.
Henceforth, Tom, Dick, and Mary will benefit from the agony I put them through. They now understand the notion of sampling as one of constructing Fourier series for periodic extensions of the "signal" in the other domain. This will allow them to meaningfully interpret their results in light of their understanding of continuous-time concepts. They know the truth.

J. R. Deller, Jr. is a professor in the Department of Electrical Engineering, Michigan State University, East Lansing, MI.

References


Editor's Note

This article has been formally reviewed by four referees according to customary IEEE procedures.

Introducing the 1994 IEEE All-Society Periodicals Package

The core collection of quality journals and magazines in electrical engineering, electronics and computing... subscribe today and get over one third of them free!

In electrical engineering, a "core" is a mass of material that directs and increases the strength of a field. The 89 journals and magazines in IEEE's All-Society Periodicals Package virtually define the fields of electrical engineering, electronics, and computing.

For a free brochure or to order call: 908-981-0060 or fax: 908-981-9667