ALL ABOUT PHASE

Robert C. Maher

Department of Electrical and Computer Engineering
Montana State University, Bozeman, MT 59717-3780 USA
rob.maher@montana.edu

ABSTRACT

While students generally demonstrate an intuitive understanding of transfer function magnitude terms such as low pass and band pass, they are often bewildered by the concepts of phase and its relationship to delay. We find that although students can easily learn the formulas and demonstrate the computer commands necessary to produce phase plots, they frequently cannot explain what the plot means, or how they might determine whether or not their implementation of a particular filter matches the theoretical phase response. To address this gap in student learning, we are experimenting with a set of lecture notes and corresponding hands-on laboratory experiments specifically dealing with system phase and delay relationships. Our preliminary results indicate that this special treatment of phase is both effective and efficient for increasing student confidence and ability in handling phase-related engineering questions.

Index Terms— Engineering education, Delay effects, Signal Processing

1. INTRODUCTION

Electrical and computer engineering students commonly first encounter the concept of *phase response* either in a linear systems and transforms course, or via Bode plot construction in an AC circuits course [1-6]. Most textbooks introduce phase through the rectangular vs. polar equivalence for expressing complex numbers or as a consequence of the properties of steady-state sinusoidal (phasor) and complex exponential analysis [1].

Although such a textbook introduction to phase is correct mathematically, the practical and physical interpretation is often left unstated, making it difficult for students to visualize relationships among time delay, phase, frequency, and waveform period. Similarly, textbook introductions to terms such as *group delay*, *phase delay*, and *minimum-phase* are typically presented as mathematical formulae without a functional, practical rationale.

We have found that students more easily grasp the important aspects of phase when they are given multiple learning opportunities: textbook description, pencil and paper calculations, computer simulation, and laboratory observation and measurement. Although our course concepts have been developed in the context of a junior/senior digital signal processing course, the material could easily be adjusted to suit the needs of a basic course in linear systems or an introductory control systems class.

1.1 An Example

A simple circuit example helps demonstrate the common misunderstandings students encounter when confronting phase plots (see Fig. 1 and Fig. 2).

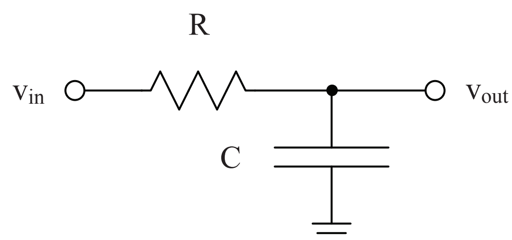

Figure 1: RC circuit example

$$H(j\omega) = \frac{1}{1 + j\omega RC} = \frac{1}{1 + \frac{j\omega}{1/RC}} = \frac{1}{1 + j\frac{\omega}{\omega_0}} \quad (1)$$

$$\angle H(j\omega) = -\arctan\left(\frac{\omega}{1/RC}\right) = -\arctan\left(\frac{\omega}{\omega_0}\right) \quad (2)$$

As educators we typically explain the procedure for constructing the Bode magnitude asymptotes as [1, 6]:

(i) flat at 0dB for $\omega<\omega_0$,
(ii) a breakpoint at $\omega=\omega_0$,
(iii) a descending asymptote of -20dB/decade for $\omega>\omega_0$.

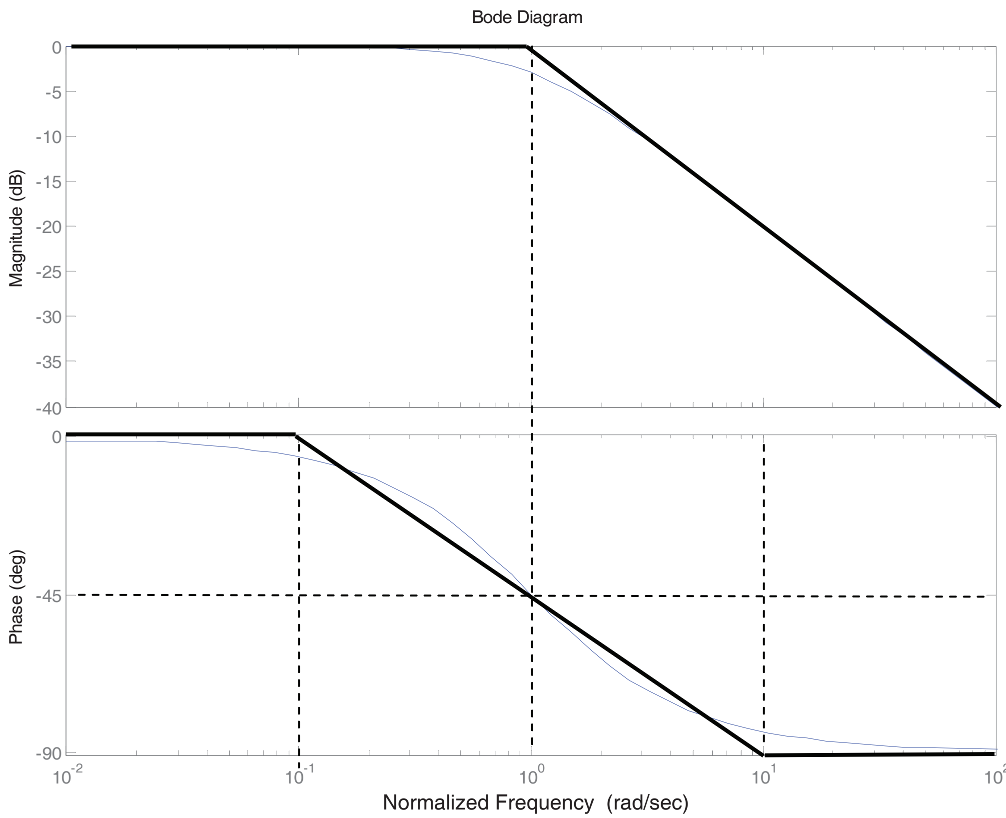

Figure 2: Bode diagram for RC circuit of Fig. 1

Similarly, for the Bode phase asymptote we teach:

(i) flat at 0 degrees for $\omega<\omega_0/10$,
(ii) a breakpoint at $\omega=\omega_0/10$,
(iii) a descending asymptote of -45°/dec for $\omega_0/10 < \omega < 10\omega_0$,
(iv) a breakpoint at $\omega=10\omega_0$,
and
(v) flat at -90° for $\omega>10\omega_0$.

Students quickly learn the step-by-step procedure for constructing such plots, and they recognize the Bode magnitude plot as being a lowpass function. In the lab, the students can assemble the RC circuit, drive it with a signal generator, and easily demonstrate that the output amplitude declines as they twist the frequency knob to higher frequencies.

The problem comes when we ask the students to describe what the phase plot *means* and how they might demonstrate this meaning in the lab. With amplitude it is relatively easy to see and verify the attenuation as frequency increases, but observing and verifying the phase can be a bit more complicated.

The remaining sections of this paper describe the features of our phase-related learning modules, including examples of the lecture, homework, and laboratory exercises.

2. CHALLENGING THE STUDENTS' PRIOR CONCEPT OF PHASE

To get started, we have the students sketch a few functions (by hand) overlapping on a single graph, such as for the range $0 < \theta < 2\pi$:

$\cos(\theta)$
$\sin(\theta)$
$\cos(\theta - \pi/2)$,

and another sketch for:

$\cos(\theta)$
$\cos(\theta - \pi/4)$.

Next, we have the students plot three cycles of each function using a computer program such as Matlab or Excel [2].

Finally, we have them modify their computer plotting routine for the case that the desired frequency is 1 kHz and we want a time range sufficient to see three cycles of the overlapping waveforms. This also requires some discussion of how best to display a continuous-time waveform as sampled data.

$\cos(2\pi(1000)t)$
$\sin(2\pi(1000)t)$
$\cos(2\pi(1000)t - \pi/2)$

and

$\cos(2\pi(1000)t)$
$\cos(2\pi(1000)t - \pi/4)$

At this point the students are invited to recognize that:

(a) $\cos(\theta)$ is ± 1 when $\sin(\theta)$ is zero, and vice versa (*which we hope they remember from prior classes!*)

(b) $\sin(\theta) = \cos(\theta - \pi/2)$ (*which they also probably remember…*)

(c) the three cycle plot of $\cos(2\pi(1000)t - \pi/4)$ *looks* just like $\cos(\theta - \pi/4)$, *except* for the labeling (scaling) of the abscissa.

(d) $\cos(\theta - \pi/4)$ is a horizontally *right shifted* version of $\cos(\theta)$: the first peak of $\cos(\theta)$ occurs at $\theta = 0$, while the first peak of $\cos(\theta - \pi/4)$ occurs *later*, namely at $\theta = \pi/4$.

Observations (a) and (b) are simply intended to be a review reminder. Observations (c) and (d) are more significant here. Observation (c) indicates that a waveform observed with an oscilloscope or a computer printout has its *phase* determined relative to the waveform reference itself, not its particular time scale or frequency, so it is often useful in a practical sense to observe the relative phase of a pair of waveforms. Observation (d) is important because it introduces the notion that a phase shift between two waveforms also represents a *time shift*.

3. ANALYTICAL PREDICTIONS VS. LABORATORY MEASUREMENTS

The analytical results for system phase obtained by interpreting the Laplace, Z, or Fourier Transforms, e.g., $\angle H(\omega)$, can give a theoretical value for the relative phase from input to output for a sinusoidal input signal of any frequency ω. When a student is given the frequency ω_0, he or she simply "looks up" the phase of the system $H(\omega)$ evaluated at $\omega = \omega_0$. However, translating this mathematical viewpoint to the lab so that the student is confident manipulating knobs on the function generator and oscilloscope takes a bit more care.

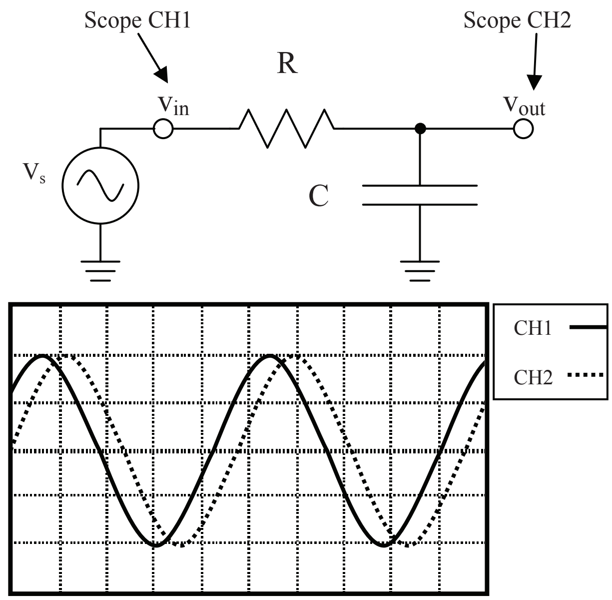

Figure 3: Lab test circuit and scope display

We have found that a useful approach is to introduce the *phase delay time* concept in the context of the lab observations. We have the students set up a simple measurement arrangement with a sinusoidal generator, simple RC network, and two-channel oscilloscope, as depicted in Fig. 3.

The key concept, of course, is for the students to understand that the phase shift is a relative quantity with respect to one whole cycle (360° or 2π radians). If the predicted phase shift at a frequency of 2 kHz is, say, π/4 radians, then the corresponding delay is expected to be π/4 out of 2π, or 1/8 of a waveform period. At 2 kHz, the waveform period, T, is (1/2000) = 500 μsec, so the expected delay is (1/8)(500 μsec) = 62.5 μsec between corresponding zero-crossings. This calculation can be expressed as:

$$\left|\frac{\textit{phase}\text{ radians}}{}\right|\frac{1\text{ period}}{2\pi\text{ radians}}\left|\frac{T\text{ seconds}}{1\text{ period}}\right| = \textit{phase delay} \qquad (3)$$

If we use the usual convention that a negative phase corresponds to a time delay, the expression can be further expressed as

$$\textit{phase delay[sec]} = -\left(\frac{1}{\omega}\right)\arg\left[H(\omega)\right], \qquad (4)$$

where $\arg[H(\omega)]$ is the principal value of $\angle H(\omega)$ [3]. Note that this also allows discussion of what is the principal value range and why it is relevant to phase calculations.

At this point, we reinforce the concepts by having the students predict the phase delay time for several different simple circuits, then while making measurements in the laboratory they complete a table of *frequency vs. phase in radians* and *phase delay in seconds* in their lab notebooks.

4. LINEAR PHASE

Based on the phase delay relationship, we ask the students what phase behavior they would expect for a system that had a constant time delay, T_{delay}, at all frequencies? In other words, a perfect delay line. Using Eqn. 4, they can see that

$$\arg[H_{lin}(\omega)] = -\omega\, T_{\text{delay}} \qquad (5)$$

From this expression, a plot of the phase $\arg[H(\omega)]$ as a function of radian frequency is a line with *slope* $-T_{\text{delay}}$, and this helps the students understand the origin and meaning of the term *linear phase* when referring to system properties. This leads naturally to discussion of the linear phase characteristics of a digital delay line N samples long, and the resulting implications for symmetrical FIR filters that also exhibit linear phase.

5. GROUP DELAY

Once the relationship between phase and time delay for steady-state sinusoidal excitation is clear, we ask the students to consider what to expect if the input signal is time-varying instead of steady-state. Specifically, the principle of *group delay* is used to help predict the delay effects of the system on the amplitude envelope of a modulated signal.

The concept is explained by having the students recognize that for a linear phase system the time delay is the negative of the *slope* of the phase curve with respect to ω:

$$-\frac{d}{d\omega}\left(\arg[H_{lin}(\omega)]\right) = T_{delay} \qquad (6)$$

For a system that does not exhibit linear phase, the time delay varies as a function of frequency (i.e., the phase slope is not a constant), but evaluating the derivative of the phase function with respect to ω at a particular frequency ω_0 can still be identified as the time delay for frequencies in the vicinity of ω_0. The phase slope approximation represents the so-called group delay of the system [3],

$$\textit{group delay of } [H(\omega)] = -\frac{d}{d\omega}\left(\arg[H(\omega)]\right) . \qquad (7)$$

6. THE TOPIC SUMMARY

The general framework described above serves as the basis for investigating phase concepts in our junior/senior level introductory DSP course. The course topics used in this particular course are summarized next.

The undergraduate DSP class begins with a review of sinusoidal signal parameters. This provides a first opportunity to probe the students' comfort with phase concepts. For example, among the first questions we pose for the students is:

> *Consider two signals: $x_1(t) = \alpha \cos(\omega_0 t)$ and $x_2(t) = \beta \cos(\omega_0 t - \pi/8)$, where α and β are assumed to be positive real constants. For $t>0$, does the first zero crossing of $x_2(t)$ occur before or after the first zero crossing of $x_1(t)$?*

Students clearly see the explicit -π/8 phase shift term in the expression for $x_2(t)$, but it often takes them a while to determine that a negative phase shift term corresponds to a right shift (delay) in time, and what the implications of this shift might be.

Once this insight is pointed out, we continue to reinforce the "sanity check" aspects of phase in subsequent exercises and homework problems, and then commence a systematic discussion of phase-related issues.

6.1 Phase measurements from analog and digital signal observation

In the lab, students make "black box" input-output observations for sinusoidal signals with a variety of frequencies, determine suitable strategies for measuring time differences, and express the waveform delay/advance in terms of radians. They also must deal with the cyclical ambiguity (wrapping) issues, e.g., is it an advance by π/4 or a delay by 7π/4?

6.2 Linear phase: the effect of a frequency-independent time delay

First in the lab and then analytically, the students learn that a constant frequency-independent time delay system corresponds to a straight line (constant slope) when the phase charcteristic is plotted as a function of frequency, and that the slope of the phase line is attributable to the system delay.

6.3 Interpretation of phase delay and group delay

Establishing the clear connection between system delay and the resulting system phase shift leads very naturally into a discussion of phase delay and group delay, as described in section 5.

6.4 Phase characteristics based on transfer function

Now that the students have a workable and practical understanding of the origin and nature of system phase, the analytical derivation of system phase from *z* or from Laplace transforms is appropriate and meaningful. The geometric approach for estimating magnitude and phase from pole-zero vectors is quite helpful, too.

6.5 Interpretation and explanation of minimum phase systems

Finally, we conclude our special treatment of phase concepts by studying the properties of minimum phase systems and help the students learn the origin and rationale of this terminology. The background is also very helpful when introducing the principles and applications of inverse filters and all-pass phase compensation filters later in the course.

7. RESULTS AND DISCUSSION

While none of the teaching strategies we employ in studying phase topics are revolutionary, we do find that the time spent on explaining and reinforcing the connections between phase response and frequency-dependent system delay is highly worthwhile.

Our assessment procedure has been informal, consisting of a simple phase pre-test at the start of the semester, followed by the special phase exercises and lab experiments. We then include specific problems on the mid term and the lab exams to help us understand whether or not the students have learned the fundamental principles. Our experience is that the students on average perform rather poorly on the pre-test despite the fact that we assume they have seen phase many times in their earlier circuits and linear systems classes. After the special exercises and reinforcement, the exam performance is substantially improved. Thus, the preliminary results indicate that this special treatment of phase is effective for increasing student confidence and ability in handling phase-related engineering questions.

We do not yet have any measurements or formal evidence regarding the level of retention for the phase-related topics emphasized in the DSP course. We plan to incorporate the phase questions into our regular curricular pre-testing arrangements that are used as part of our ongoing accreditation assessment.

8. REFERENCES

[1] Hambley, A.R., *Electrical Engineering Principles and Applications*. 4th ed. Pearson Education, Inc., Upper Saddle River, New Jersey, 2008.

[2] McClellan, J.H., R.W. Schafer, and M.A. Yoder, *DSP FIRST: A Multimedia Approach*. Prentice Hall, Upper Saddle River, New Jersey, 1998.

[3] Oppenheim, A.V., R.W. Shafer, and J.R. Buck, *Discrete-Time Signal Processing*. 2nd ed. Prentice Hall, Upper Saddle River, New Jersey, 1999.

[4] Orfanidis, S.J., *Introduction to Signal Processing*. Prentice Hall, Englewood Cliffs, New Jersey, 1996.

[5] Proakis, J.G., and D.G. Manolakis, *Digital Signal Processing*. 4th ed. Pearson Prentice Hall, Upper Saddle River, New Jersey, 2007.

[6] Sedra, A.S., and K.C. Smith, *Microelectronic Circuits*. 5th ed. Oxford University Press, USA, 2007.