[**DOING PHYSICS WITH MATLAB**](http://www.physics.usyd.edu.au/teach_res/mp/mphome.htm)

**WAVE MOTION**

# 

# THE [1D] SCALAR WAVE EQUATION

# THE FINITE DIFFERENCE TIME DOMAIN METHOD

# 

# 

Ian Cooper

School of Physics, University of Sydney

ian.cooper@sydney.edu.au

[**DOWNLOAD DIRECTORY FOR MATLAB SCRIPTS**](http://www.physics.usyd.edu.au/teach_res/mp/mscripts)

The following mscripts are used to solve the scalar wave equation using the finite difference time development method. All the mscripts are essentially the same code except for differences in the initial conditions and boundary conditions. Graphical outputs and animations are produced for the solutions of the scalar wave equation.

**em\_swe\_03.m**

models the propagation of either a rectangular pulse or a Gaussian pulse

**em\_swe\_04.m**

models the superposition and interference of propagating waves

**em\_swe\_05.m**

models the propagation of traveling harmonic waves

**em\_swe\_06.m**

models standing waves

**em\_swe\_07.m**

models the reflection and transmission of a Gaussian pulse at an interface

**Introduction**

We will use the **finite difference time domain (FDTD) method** to find solutions of the most fundamental partial differential equation that describes wave motion, the **one-dimensional scalar wave equation.**

The [1D] scalar wave equation for waves propagating along the X axis can be expressed as

(1) 

where  is the **wavefunction** and *v* is the speed of propagation of the waveform. For electromagnetic waves, the wave function represents the electric field or magnetic field; for sound waves, the wave function represents the pressure or particle displacement fluctuations; and for waves on strings, the wave function gives the string displacement.

**Finite Difference time Development Method**

The FDTD method can be used to solve the [1D] scalar wave equation. In the numerical solution, the wavefunction is approximated at discrete times and discrete grid positions. Let the execution time for a simulation be given by *T*. The time step is and the number of time steps is *Nt.* The wavefunction  is calculate at *Nx* discrete positions from *x* = 0 to *x*= *L* where the spacing between the grid points is 

The *ntt*h time interval is



The *nx*th grid position is



The wavefunction  at time and grid position  is expressed in terms of the time index *nt* and the spatial index *nx* as 

The second partial derivative of the wavefunction in space at time  for  is approximated by the central difference formula

(2) 

The central difference approximation to the second partial time derivative of  for  at the grid position  is

(3) 

Substituting the two central difference expressions of equations 2 and 3 into the [1D] scalar wave equation (1), we get for the latest value of the wavefunction at the grid point 

(4) 

This is a fully explicit second-order accurate expression for  in that all wave quantities on the right hand side are known, that is, they were obtained during previous time steps, *nt* and .

The wavefunction values have to been assigned at each grid point for the first two time steps to start the solution process given by equation 4





Using the assigned values for the first two time steps and the specified boundary conditions, equation 4 is repeated used for all space points to calculate the solution of the scalar wave equation at all later times.

A fundamental assumption in the FDTD method is that in the calculation of the wavefunction at a given grid point we need to know the two adjacent values of the wavefunction, but at the ends of the grid, this is not possible. To solve the scalar wave equation, the wavefunction at the grid points  and  must be given at all times, that is, the **boundary conditions** must be specified. There are several different types of boundary conditions:

**Dirichlet** *u(*boundary) = 0 **Fixed end**

**Neumann**  **Free end**

Implementing this boundary condition at  and at  using a forward difference approximation for the derivative gives



**Absorbing boundary condition** (**ABC**): This is an important type of boundary condition, for example, most electromagnetics problems involve unbounded regions, which cannot be modelled computationally. One option is to use one of the above boundary conditions and make the simulation region very large, and terminate the simulation before reflections from the boundary perturb the solution in the region of interest. The drawback of this approach is that the larger the simulation region, the greater the computational cost of the simulation. A better approach is to use a boundary condition that absorbs waves and reflects as little energy as possible.

**Numerical Stability and the Magic Time Step**



The accuracy of the FDTD method is very dependent upon the time step  size and the size of the grid spacing . The grid space resolution directly impact the numerical phase velocity. It is useful to define the numerical stability factor called the **Courant number** *S* where

(5) 

and the **grid sampling resolution** 

(6) 

 is the “real-world” wavelength and not the numerical wavelength in the FDTD grid. To have numerical stability the condition must be satisfied.

For a particular choice of grid  the corresponding spatial sampling is



where is the maximum sampling rate and  the minimum free-space wavelength that can be sampled without aliasing. Hence,



and since  then .

These conditions imply that in one time step  a wave can propagate at most one grid point .

Of particular interest is the case where  or  and the results of using this time step is so remarkable that it is called the **magic time step**. When the magic time step is used the numerical solution to the scalar wave equation is an **exact solution** and not an approximation. We can proceed any number of time steps of the FDTD algorithm, always calculating the exact propagating waveform.

**Rectangular Pulse Propagation**

We will give a number of example calculations of the propagation of pulses in a *Nx* = 200 grid point range of length *L* = 100 a.u. and grid spacing *hx =* 0.500 a.u. The free space propagation velocity is set to *v*= 10 a.u.

Running the simulations for different Courant numbers *S* for a rectangular pulse, clearly demonstrates, that in the solving the [1D] scalar wave equation it is best to use the magic time step where *S* = 1.

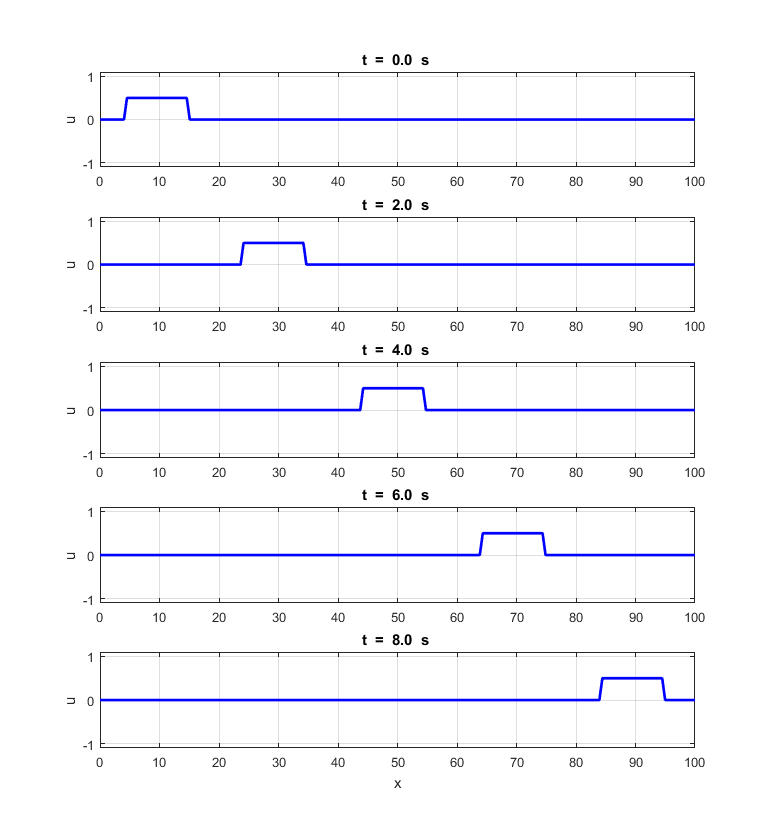
**Rectangular pulse *S* = 1 magic time-step**

Fig. 1. Position of the rectangular pulse at fixed time intervals for *S =* 1 (magic time-step). *Nt* = 160 **em\_swe\_03.m**

Figure 1 shows the position of the pulse after fixed time intervals when the time step is set equal to the magic time step where the Courant number is *S* = 1. The shape and spatial width of the pulse are completely preserved as it propagates along the X axis. The free space velocity is  a.u. The numerical velocity *vn* can be measured from the displacement of the pulse *dx* from *t* = 0 a.u. and at the end of the execution time *dt* = 8.00 a.u. The uncertainly in the measurement of the displacement is .

 a.u. *measured using Matlab Data Cursor tool*

The free space velocity and the numerical velocity of propagation are the same when *S* = 1 (the time step is set equal to the magic time step).

**Rectangular pulse *S* = 0.95**

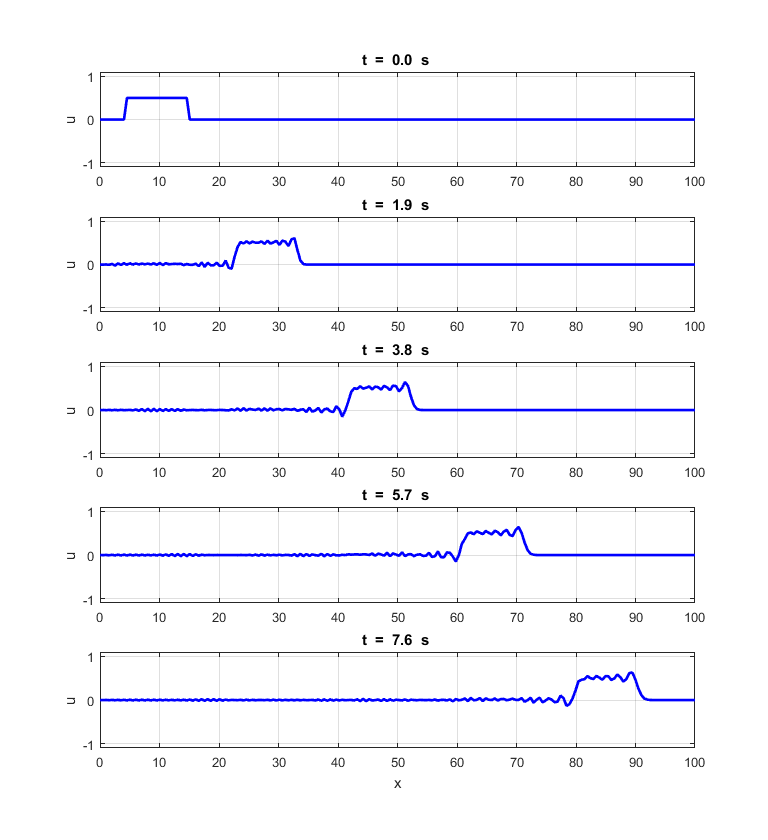


Fig. 2. Position of the rectangular pulse at fixed time intervals for *S =* 0.95. *Nt* = 168 **em\_swe\_03.m**

Figure 2 shows the propagation of the pulse when *S =* 0.95 < 1. There is a ringing effect at the leading and trailing edges of the pulse. The numerical velocity is . The ringing represents time retarded propagation of the sparsely, high frequency spectral components of the discontinuities at the steps of the rectangular pulse.

**Rectangular pulse *S* = 0.70**

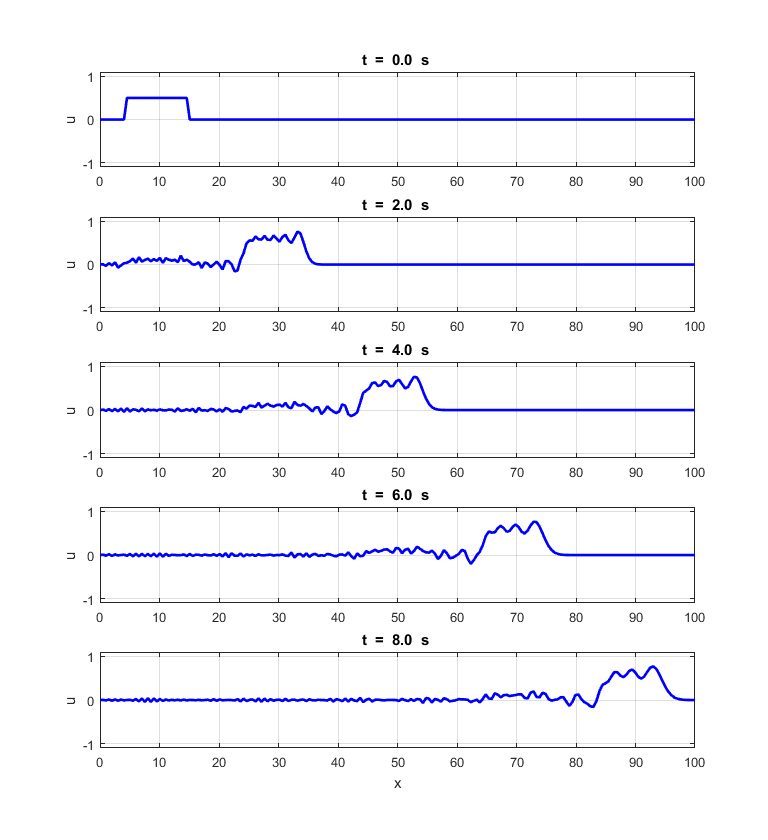


Fig. 3. Position of the rectangular pulse at fixed time intervals for *S =* 0.70. *Nt* = 230 **em\_swe\_03.m**

If you carefully compare figures 1 and 3, there is a weak **superluminal response** () that propagates just ahead of the leading edge of the rectangular pulse. After the time interval of *dt =* 8.00 a.u. the leading edge of the pulse for *S* = 1.00 is at *x * 95 a.u. whereas for the pulse for *S*= 0.7, the position of the leading edge is at x  100 a.u.

In the numerical solution for  some of the spectral components of the pulse travel at a speed slower than *v* whilst other spectral components travel faster than *v.* As well as the ringing effect, there is broadening of the pulse as it propagates.

**Gaussian pulse propagation**

We can study the propagation of pulses which have an initial Gaussian spatial profile using the magic time step. For the magic time step, the pulse advances one grid space in one time step.

The Gaussian pulse is specified in the Matlab mscript as a function of the spatial index *nx.* The input parameters for the Gaussian pulse are

% Gaussian pulse

A = 0.8; % pulse amplitude

s = 5; % pulse width

np = 20; % pulse starting point given by index np

Gaussian profile is specified at time steps 1 and 2

if flag == 1 % Gaussian pulse: 1st & 2nd time steps

for nx = 2 : Nx-1 % time step nt = 1

u(nx,1) = A .\* exp(-0.5 .\* ((nx-np)./s).^2);

end

for nx = 2 : Nx - 1 % time step 2

u(nx,2) = A .\* exp(-0.5 .\* ((nx-np-1)./s).^2);

end

end

When the magic time step is used the Gaussian profile and speed of propagation are preserved as shown in figure 4.

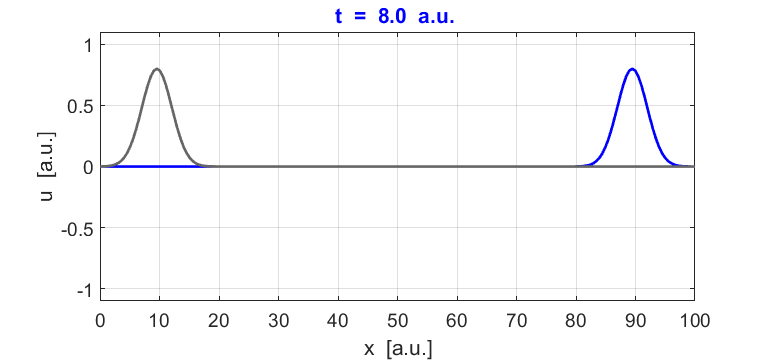


Fig. 4. Propagation of a Gaussian pulse. The pulse starts at position *x*= 10 a.u.at time *t* = 0 and at finishes at position *x*= 90 a.u. after a simulation time of *t* = 8.0 a.u. **em\_swe\_03.m**

The numerical speed of the pulse can be calculated in the Matlab Command Window using logical functions

 *dx* = distance travelled in the time interval *dt*

>> (x(u(:,end) == max(u(:,end))) - x(u(:,1) == max(u(:,1))))/8

ans = 9.9874

The free space wave speed is *v* = 10 a.u. and is its value agrees with the numerical wave speed *vn* = (9.99 ± 0.02) a.u.

**Boundary Conditions**

We can use the Gaussian pulse to investigate the behaviour of the pulse for different boundary conditions. The same boundary conditions apply at each time step for *nt* = 1 to *nt* = *Nt -*1

**Dirichlet** boundary conditions *u(*boundary) = 0

u(1, nt+1) = 0; % Fixed end

u(Nx, nt+1) = 0; % Fixed end

**Neumann**  boundary conditions 

u(Nx,nt+1) = u(Nx-1,nt+1); % Free end

u(1,nt+1) = u(2,nt+1); % Free end

**Absorbing boundary conditions ABC**

Using the magic time step, it takes one time step for the wave to propagate from one grid position to the next. Hence, to apply absorbing boundary conditions at the ends of the simulation space, the values of the wavefunction at the boundaries are set to the adjacent wavefunction values one time step earlier.

u(Nx,nt+1) = u(Nx-1,nt); % ABC

u(1,nt+1) = u(2,nt); % ABC

Figures 5 and 6 show the graphical output for a number of simulations of a pulse being reflected from the boundaries of the simulation region.

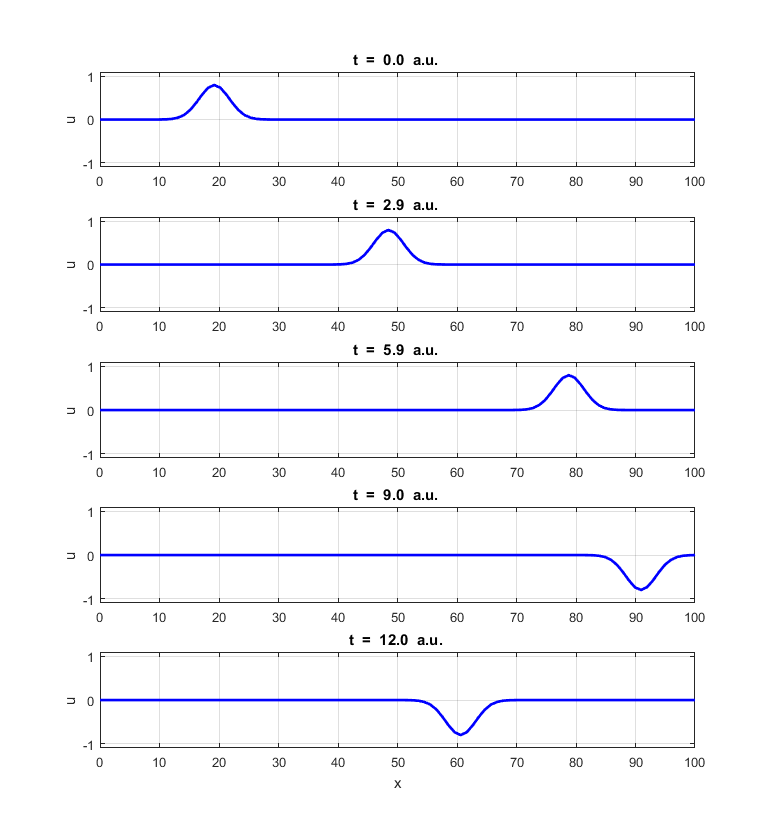


Fig. 5. A Gaussian pulse reflected a fixed end termination

*u(*boundary) = 0. The reflected pulse is inverted because of a change in phase upon reflection. **em\_swe\_03.m**

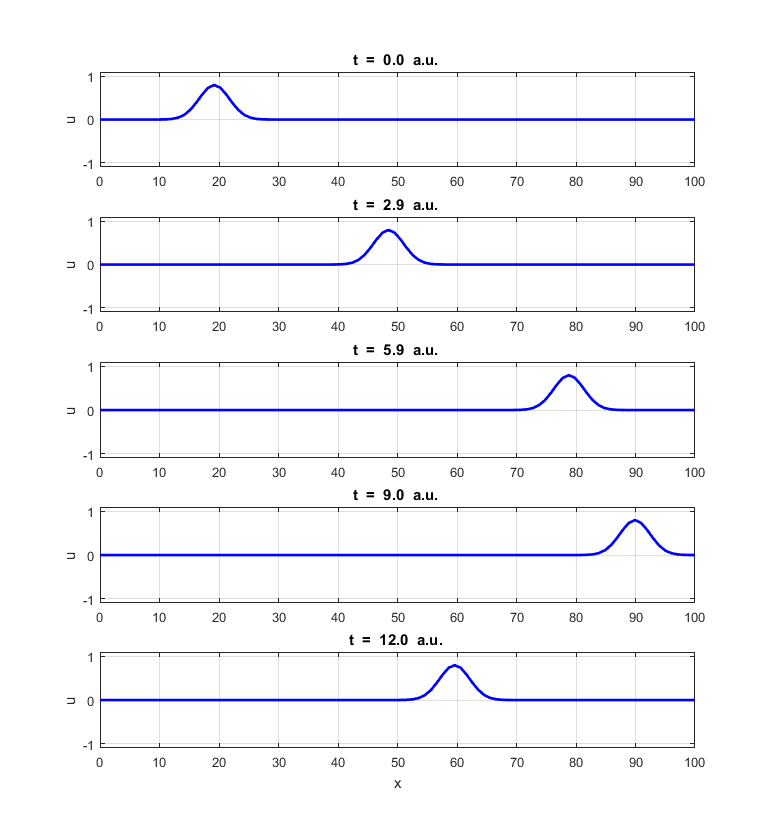


Fig. 6. A Gaussian pulse reflected a free end termination

. The reflected pulse is upright because of a zero change in phase upon reflection. **em\_swe\_03.m**

You can view a simulation of a pulse of being reflected from the end *x*= *L* which is a free end and from the end *x* = 0 which is a fixed end.

[view animation](http://www.physics.usyd.edu.au/teach_res/mp/images/ag_swe_001.gif)

**Superposition Principle: Interference**

We can model the superposition of waves by considering two pulses moving in opposite directions. If one pulse is given by the wavefunction  and the other pulse by and both wavefunctions are solutions of the scalar wave equation (equation 1) then using the property of **linearity**, the sum of any number of solutions is also a solution. This fact is called the **superposition principle**. Hence, when two or waves combine, the resultant wave is the algebraic sum of the individual waves.

The total function  for our two pulses is simply the sum of the two individual wavefunctions



The superposition principle vastly increases the number of waveforms that waves can take and gives rise to a rich palette of patterns called **interference**. When the two pulses add to give a maximum amplitude of the result waveform, the waves are said to **interfere constructively**. **Destructive interference** occurs when the amplitude of the waves cancel each other giving a minimum amplitude of the resultant waveform.

Figure 7 shows the two pulses interfering constructively as they pass through each other and figure 8 shows destructive interference. Notice that the pulse simply pass through each other.

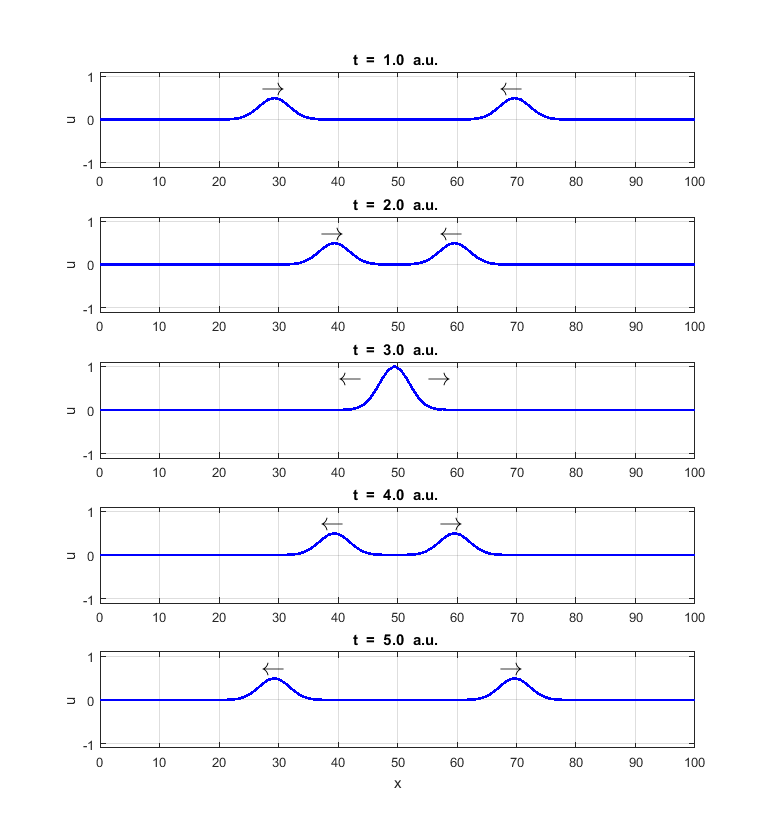


Fig. 7. Two pulses traveling in opposite directions interfere with each other. At time *t* = 3.0 a.u. the pulses interfere constructively, the resultant waveform has a maximum amplitude. **em\_swe\_04.m**

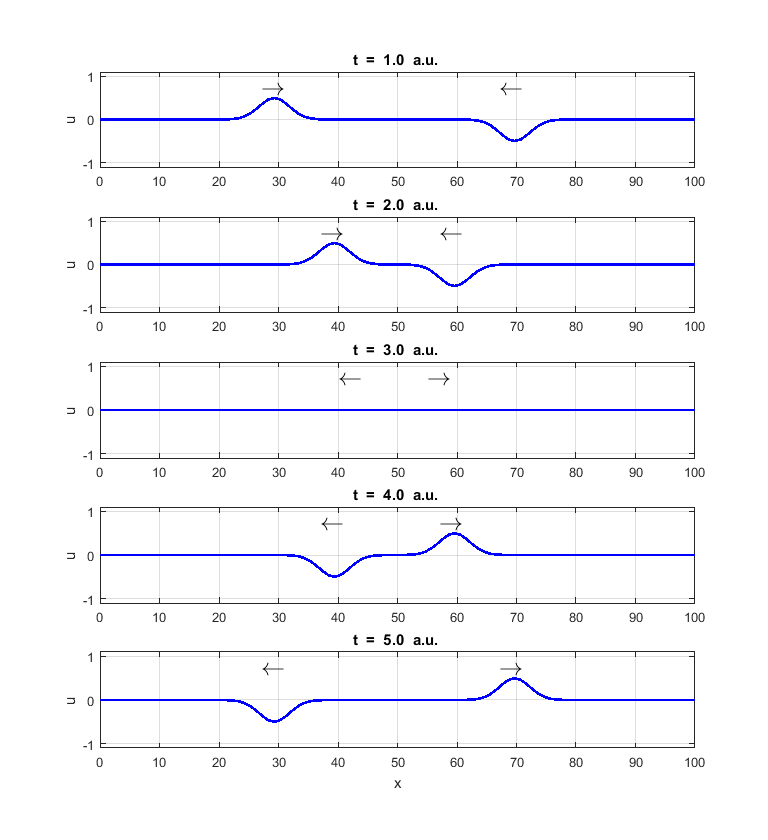


Fig. 8. Two pulses travelling in opposite directions interfere with each other. At time *t* = 3.0 a.u. the pulses interfere destructively and the resultant waveform has a minimum amplitude. **em\_swe\_04.m**

The wavefunction  is a function of both position and time. Figures 7 and 8 shows the position at the pulse at fixed times. However, Matlab can be used to create animated gifs. You can view an animation of colliding pulses which exhibit both constructive and destructive interference. The animation gives you a better picture of the interference phenomenon.

[view animation](http://www.physics.usyd.edu.au/teach_res/mp/images/ag_swe_002.gif)

**Traveling waves**

We can model harmonic waves propagating along the simulation space. At *x* = 0, a source point is used to set up the traveling wave which moves from the left to the right

 sinusoidal source at *x*  = 0

At the end of the simulation region, , we need to employ absorbing boundary conditions, otherwise, a reflected wave traveling from right to left would interfere with the incident waveform traveling left to right. Once the sinusoidal wave has been established, each grid point moves up and down with simple harmonic motion and the speed of the wave is the free space speed *v.*

Figure 9 shows the waveform at fixed time intervals for a wave with the source parameters *A*= 1, *T =* 2.5 a.u. and *v* = 10 a.u. The theoretical wavelength of the traveling wave is  From figure 9 you can estimate the speed of the wave, the wavelength and then calculate the period of the wave using the Matlab Data Cursor tool

Numerical velocity 

Numerical wavelength 

Numerical period 

The numerical values agree well with the specified values within the uncertainties in making the measurements on the course grid.

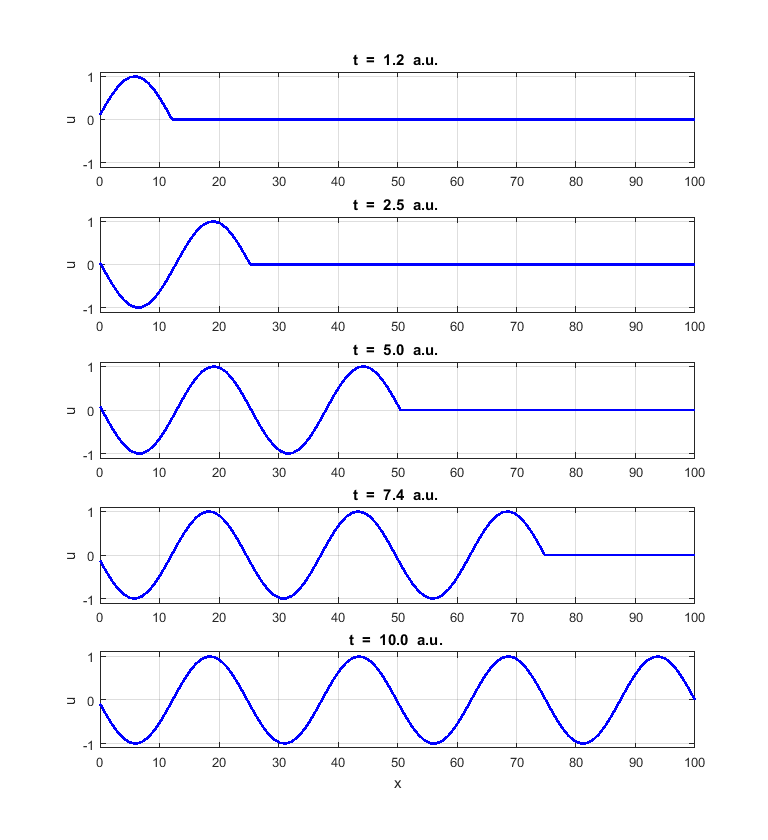


Fig. 9. Harmonic wave at fixed time intervals due to a sinusoidal point source located at *x* = 0. Absorbing boundary conditions are employed at *x = L* so that there is no reflected wave at the boundary. **em\_swe\_05.m**

[view animation](http://www.physics.usyd.edu.au/teach_res/mp/images/ag_swe_005.gif)

**Standing Waves**

We can model the behaviour of waves confined in [1D] space by finding the wavefunction that is a solution to the scalar wave equation. In our model no energy is dissipated from the vibrating system. We start with the wavefunction being zero at all grid point  and a small amplitude sinusoidal point source applied to the boundary at  (origin). The wavefunction is set to zero at the boundary .

Boundary conditions

.

 sinusoidal source at 

A node occurs when the value of the wavefunction is zero at all times. A true node exists in our model at  and provided the amplitude *A* is small than at  we can assume the boundary condition approximates a node and behaves as a fixed end. Therefore, the wave will be confined to the region from  to  since reflections occur at both boundaries.

For our model we will use the input parameters (all units are arbitrary)

Length of confined space *L* = 100

Number of grid points *Nx* = 200

Number of time steps *Nt*  variable

Speed of wave *v* = 10

Courant number *S* = 1 (use magic time step)

Sinusoidal point source located at the origin

Amplitude A = 0.1

Period *T*  variable

Boundary condition at  is  (fixed end)

Wavelength 

The sinusoidal vibration of the source at the origin initiates a disturbance that travels at a speed *v* from the left to the right towards the boundary at . When the traveling sinusoidal wave reaches this boundary, the wave is reflected. The wave is then reflected from the boundary at the origin. The waves are now confined in space with a node at each end and travel left to right and right to left interfering with each other according to the superposition principle. The sinusoidal source at  is continually adding energy to the wave and energy is returned to the source from the wave moving from right to left. A wave is built-up that vibrates at the frequency of the source. At most frequencies very little energy is added from the source to the wave.

However, at certain frequencies of vibration of the source, energy is very efficiently added to the wave and large amplitude vibration are created and give a stationary vibration pattern called a **stationary wave** or **standing wave**. The frequencies at which the confined wave can vibrate at are called the **natural frequencies** of vibration. When a system is disturbed, it tends to vibrate in a mode that is a mixture of its natural frequencies of vibration. The source frequency is called the **driving frequency**. When the driving frequency matches a natural frequency, large amplitude vibrations can occur. This phenomenon is called **resonance** and the frequencies are often referred to as the **resonance frequency**.

The time taken for the disturbance to be travel from the origin and return is . Just as the disturbance returns to the origin, the source produces a disturbance that reinforces the wave and energy is therefore efficiently transferred from the source to the wave. If the source vibrates sinusoidally with a period equal to  then energy will be continually added to the wave building up its amplitude. The natural frequency for this mode of vibration is called the **first harmonic** or the **fundamental frequency**.



If the source driving frequency  then the rate at which energy is transferred from the source to the wave is maximized and large amplitude vibrations will occur. ***n***is referred to as the **mode number**, for example, if , then  is called the **2nd harmonic**.

When a stationary waveform pattern is established in the confined space with a node at each end , the waveform is characterised by a sequence of equally spaced nodes ( zero amplitude) and antinodes ( maximum amplitude) at fixed positions in the confined space. The positons of the nodes and antinodes do not vary with time. The spacing between adjacent nodes or adjacent antinodes is equal half-wavelength



For our specified parameters the resonances modes for the wave confined in space for the first five harmonics are



Figure 10 shows the time evolution of the wavefunction when the source frequency is not equal to a natural frequency. Only a small amplitude wave is created and there are no nodes or antinodes.

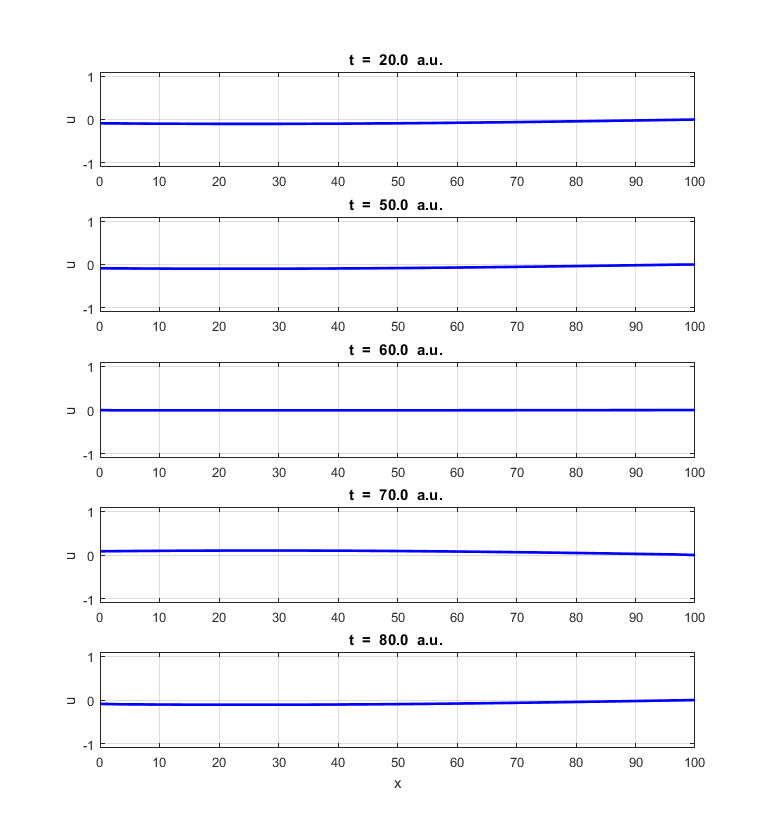


Fig. 10. The driving frequency of the source is not equal to a natural frequency. Only a small amplitude wave is created that oscillates at the driving frequency. **em\_swe\_06.m**

Figures 11 to 15 shows the time evolution of the wavefunction as a function of position in the confined space for the first five harmonics. The amplitude of the wave increases with time as energy is continually transferred from the source to the wave. For each harmonic that is excited, a standing wave pattern is established with the sequence node, antinode, node, … , antinode, node. The location of the nodes and antinodes varies slightly with time because there is not a true node at the origin ().

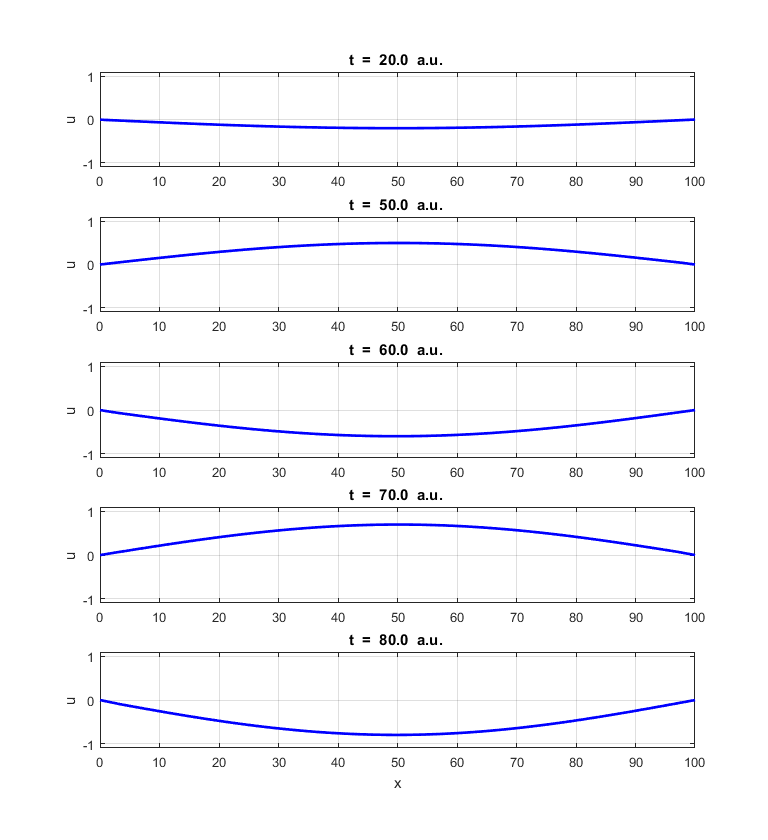


Fig. 11. Fundamental or 1st harmonic mode.

 **em\_swe\_06.m**

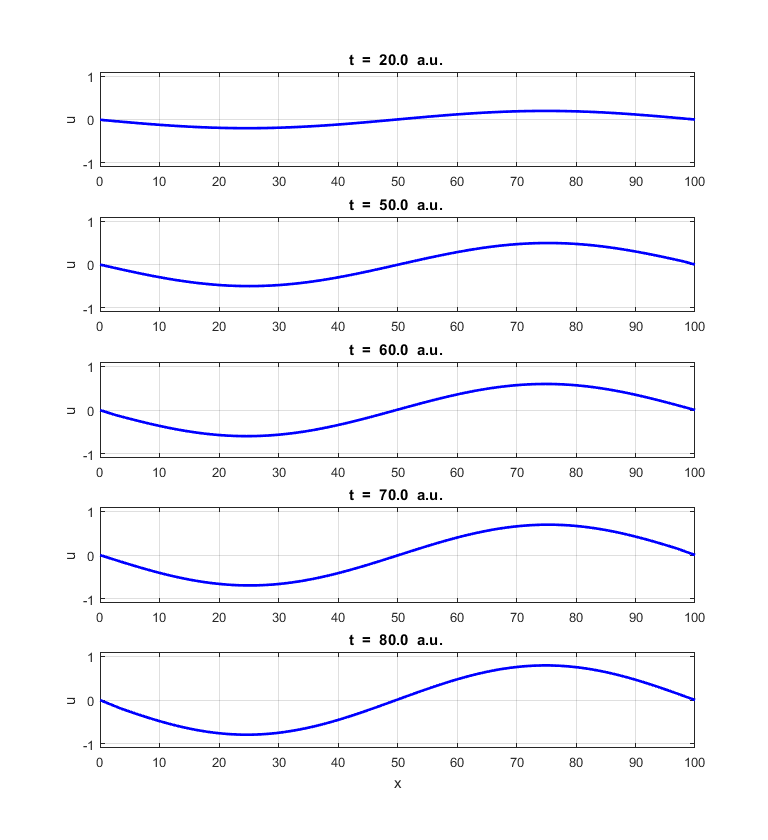


Fig. 12. 2nd harmonic mode.

 **em\_swe\_06.m**

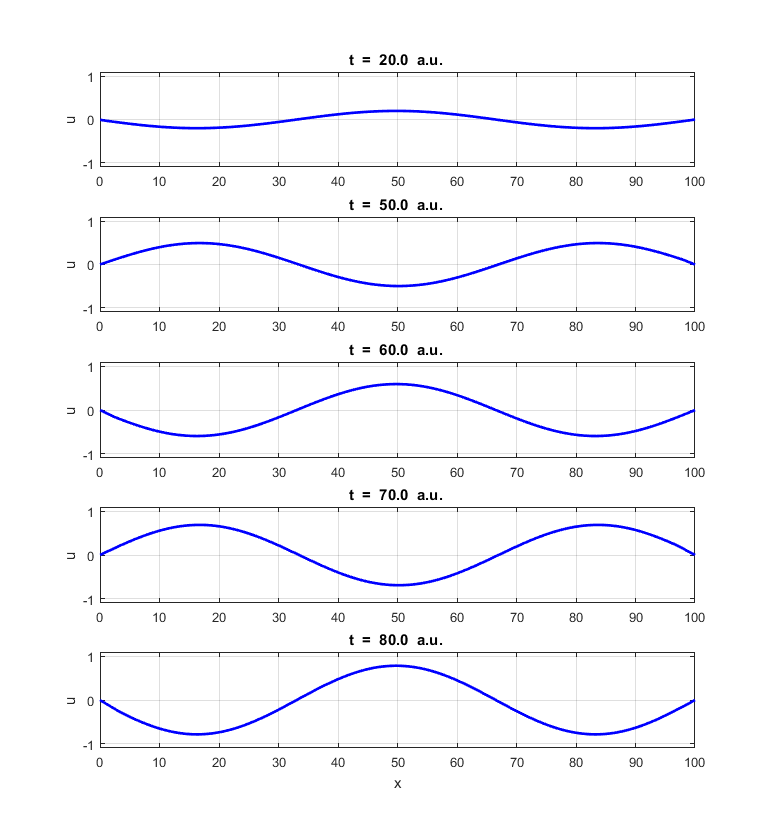


Fig. 13. 3rd harmonic mode.

 **em\_swe\_06.m**

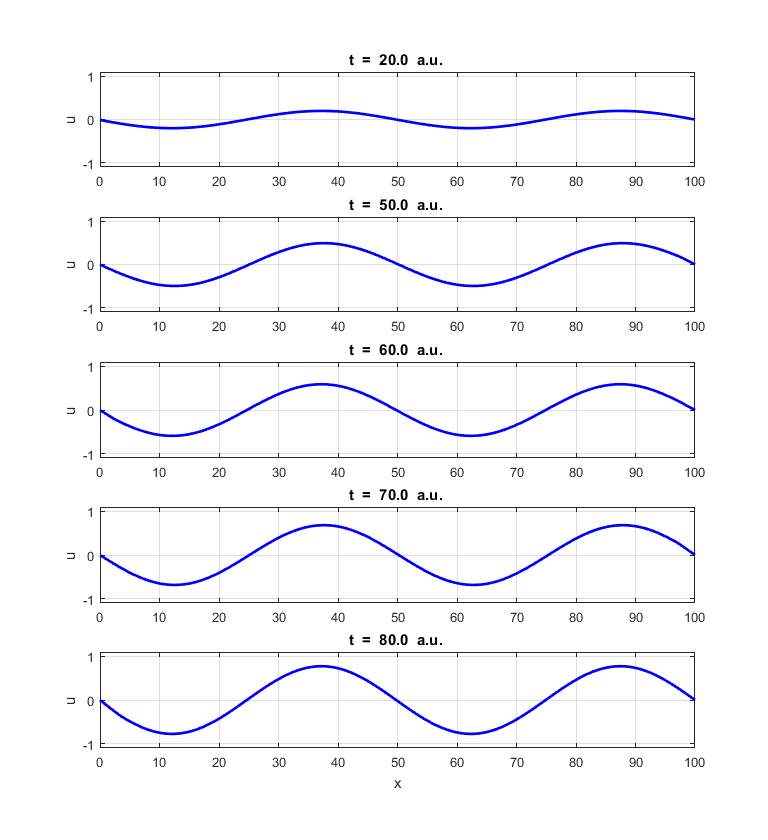


Fig. 14. 4th harmonic mode.

 **em\_swe\_06.m**

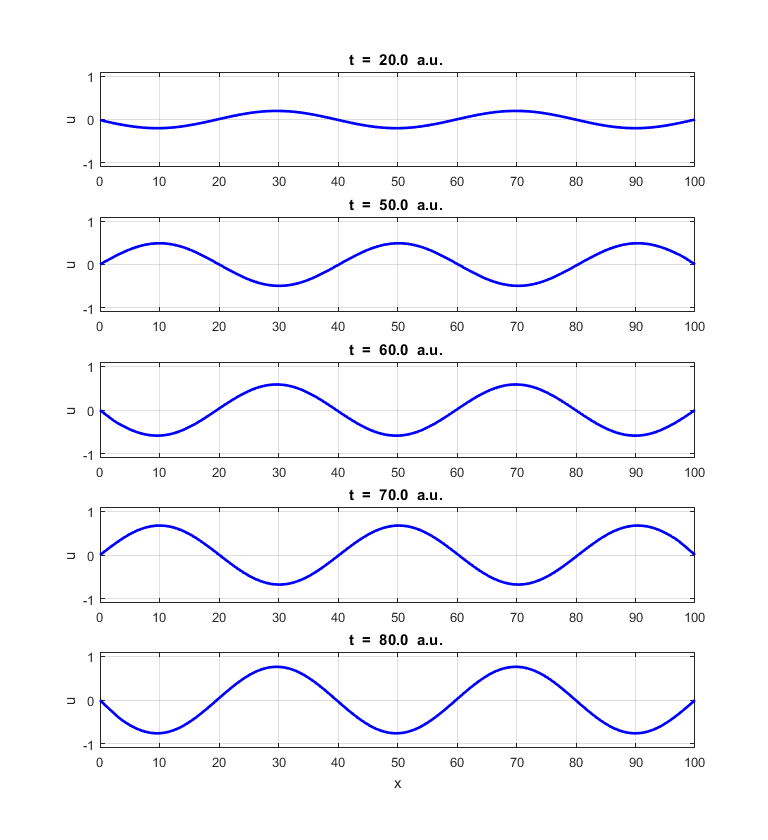


Fig. 15. 5th harmonic mode.

 **em\_swe\_06.m**

We can change the boundary conditions so there is always an antinode at . The wave is still confined in the region from  to  because of the reflections at the boundaries. When the source frequency matches a natural frequency, large amplitude oscillations are created.

**Boundary conditions**





**Normal modes** 



 Fundamental or 1st harmonic

Only the odd harmonics are natural frequencies 



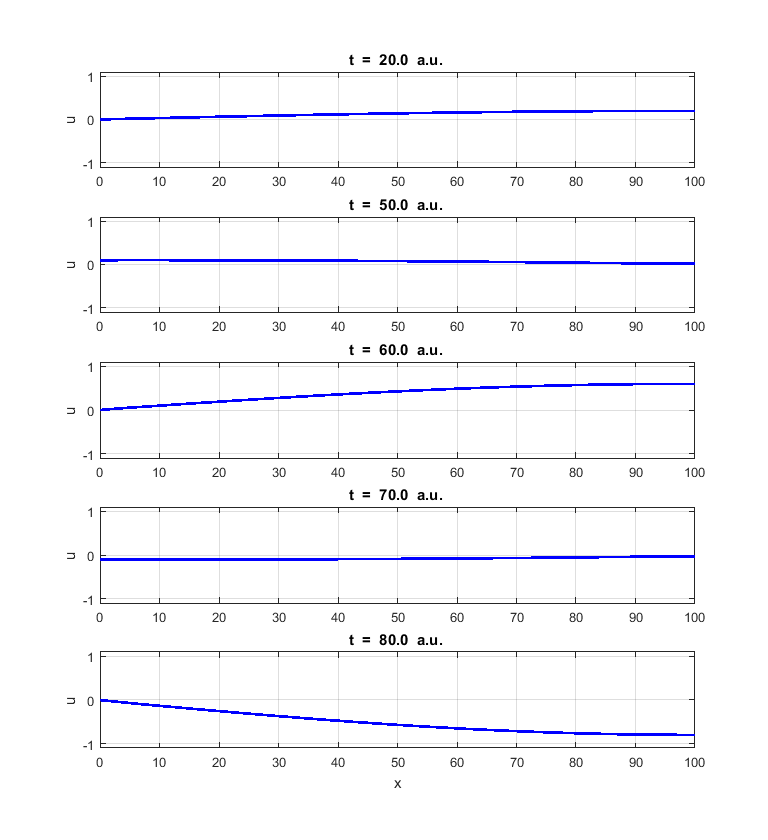


Fig. 16. Fundamental or 1st harmonic

 **em\_swe\_06.m**

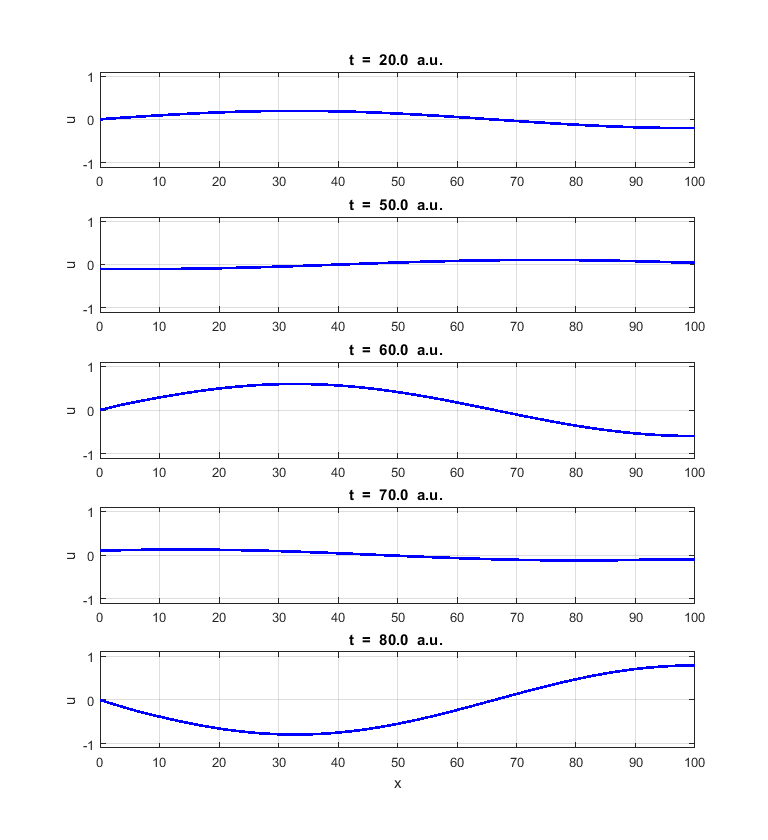


Fig. 17. 3rd harmonic



**em\_swe\_06.m**

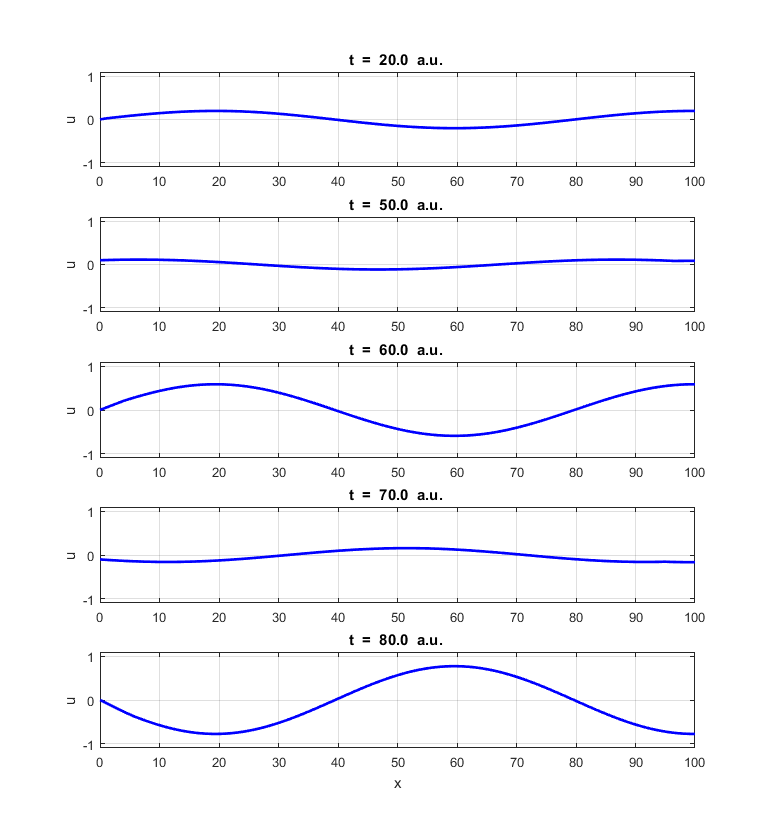


Fig. 18. 5th harmonic



**em\_swe\_06.m**

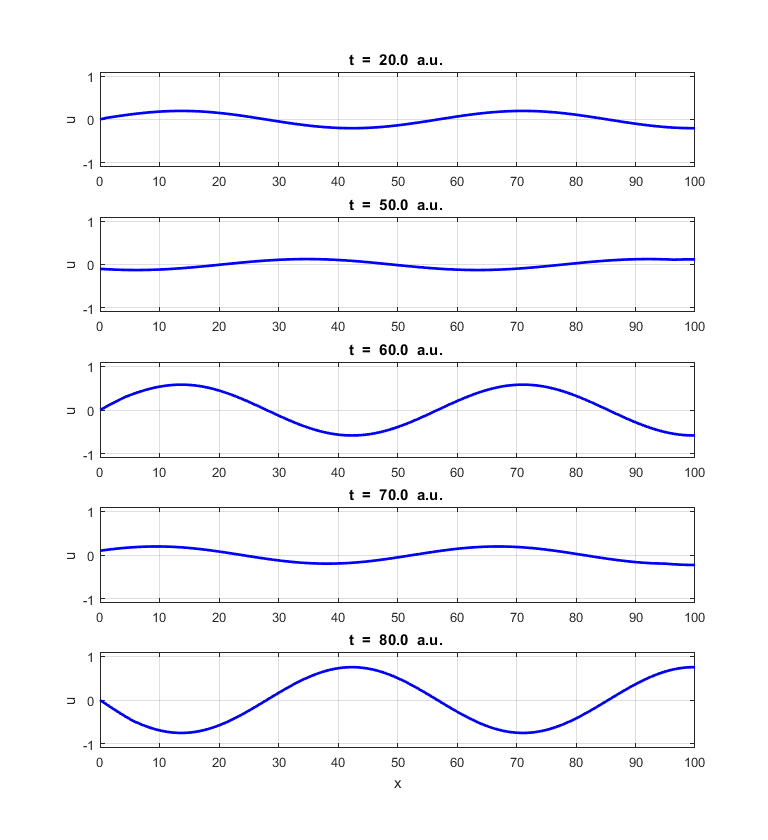


Fig. 19. 7th harmonic



**em\_swe\_06.m**

[view animations of standing waves](http://www.physics.usyd.edu.au/teach_res/mp/doc/em_swe_standing.htm)

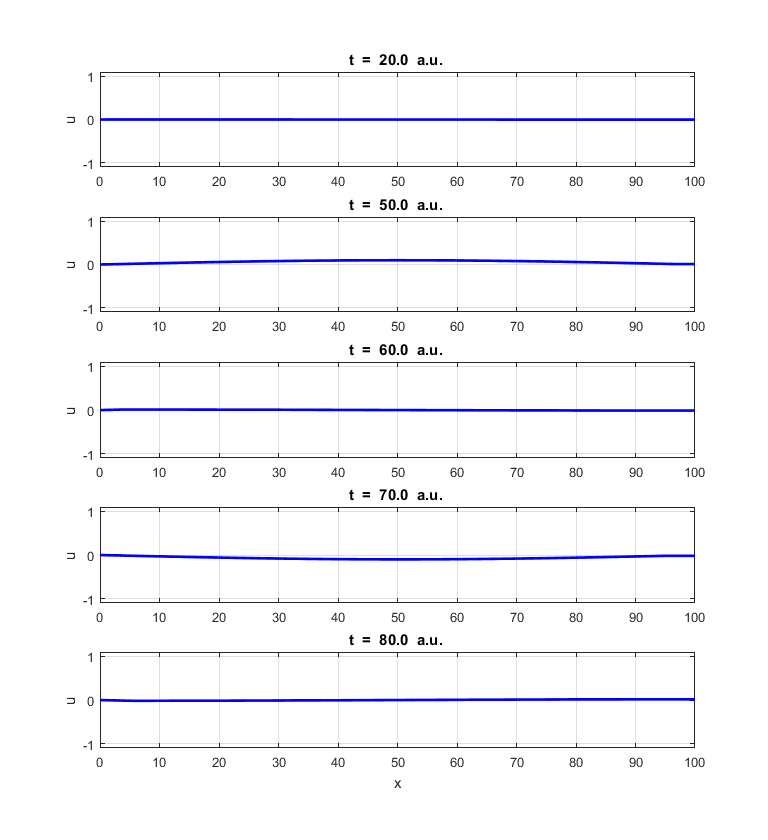


Fig. 20. Source driving frequency is equal to the frequency of the second harmonic and this does not equal one of the natural frequencies. A large amplitude standing wave is not created.

**em\_swe\_06.m**

**Reflection and Transmission at an interface**

Waves experiences partial transmission and partial refection at an interface when there is a difference in velocity of the wave on either side of the interface. Reflection is a wave phenomenon that changes the direction of a wavefront at the interface. Transmission occurs when some of the energy of the incident wave is transferred across the interface.

We can define the reflection coefficient *r* and the transmission coefficient *t* as



where  is the amplitude of the incident wave and *v*1 is the speed of the incident wave,  is the amplitude of the reflected wave travelling at speed *v*1and  is the amplitude of the transmitted wave traveling at a speed *v*2*.*

We can model the reflection and transmission of waves at an interface by solving the scalar wave equation in which the wave speed to the left of the interface is *v*1 and *v*2 to the right of the interface. If the interface position is  then



Figure 21 shows the propagation of a Gaussian pulse striking the interface where . The incident pulse has an amplitude  traveling from left to right and then striking the interface. There is excellent agreement between the theoretical and simulation values for the amplitudes of the reflected and transmitted pulses:

Theory 

Simulation 

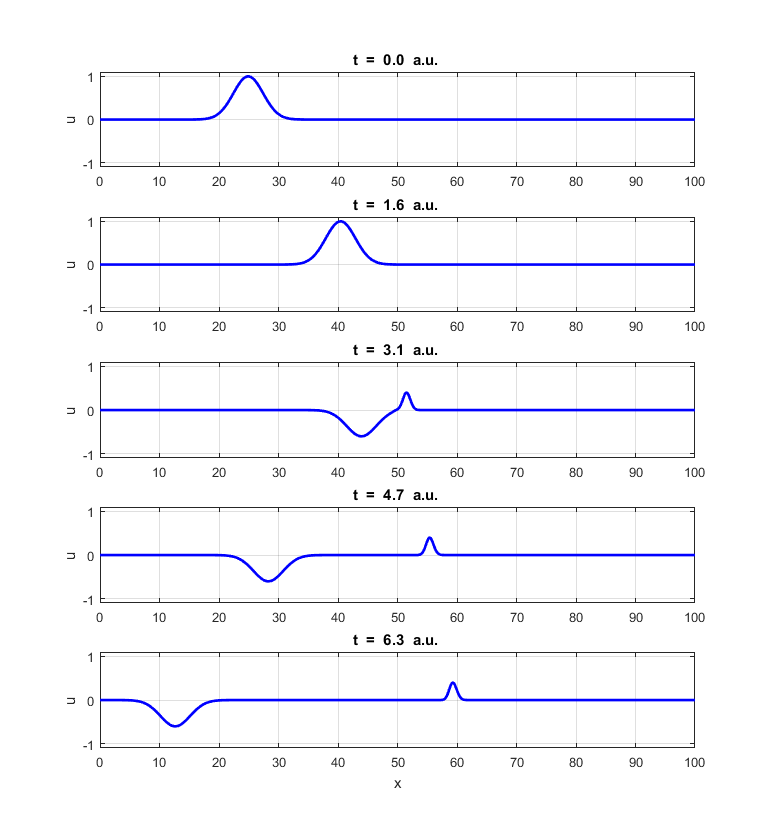
The incident pulse has a width of  between the  points. The width of the reflected pulse is which is almost the same as the width of the incident pulse. The spatial width of the transmitted pulse is  which is about ¼ of the incident pulse width as predicted by the analytical theory. 

Fig. 21. Time evolution of a Gaussian pulse striking an interface at 



[view animation](http://www.physics.usyd.edu.au/teach_res/mp/images/ag_swe_07_1.gif)

Both regions of the grid are modelled with the same values for  and  but the Courant numbers are different  and . We are no longer using the magic time step, therefore, our solution is only an approximation. In figure 21 the number of space grid point is . If we run the simulation with as shown in figure 22, then the results of the simulation are not accurate. Generally, the smaller the value assigned for , the more accurate the numerical predictions.

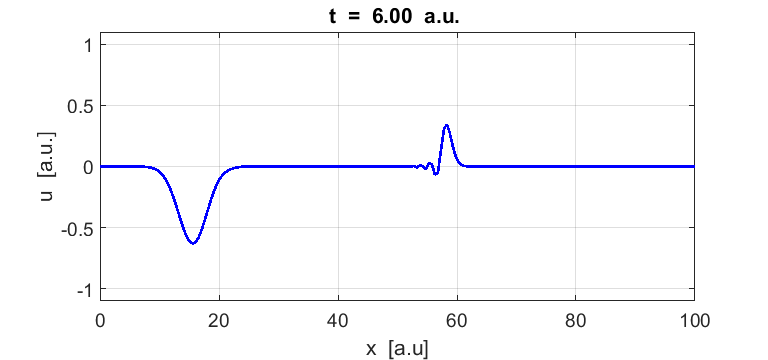


Fig. 22. Simulation showing a numerical instability in the solution. Generally, the smaller the values for the spatial and time increments the better the numerical stability and the more accurate predictions.