Emergence in Action: Particles as Field Structures

Part 2: From Spring Mechanics to a Relativistic Field

I. I. The Harmonic Oscillator

I.1 An Ideal Spring

Take a spring. An ordinary spring. Fix it to the ground and attach a small mass to its upper end. The mass has an equilibrium position. In that position no net force acts on it; if it is initially at rest, it therefore remains at rest. Pull it away from equilibrium and a restoring force, produced by the elasticity of the spring, appears. If the mass is released, this force accelerates it towards the equilibrium point. Because of its inertia, however, it does not stop there. It continues beyond equilibrium until it has travelled approximately the same distance on the other side. Once the mass has crossed equilibrium, the restoring force points upwards. The mass eventually reverses direction and continues moving back and forth until friction dissipates its mechanical energy as heat and it returns to equilibrium.

Here, however, we are not describing real life in all its complexity. We are constructing a model and making the following simplification: there is no thermal dissipation. The mass therefore continues oscillating forever. We shall make another idealisation as well: the experiment takes place in space, far from any significant gravitational source, so we need not consider the weight of the small mass. To ensure that the spring remains perfectly vertical, we may imagine it mounted on a rigid guide rod.

I.2 Dynamics of a Free Harmonic Oscillator

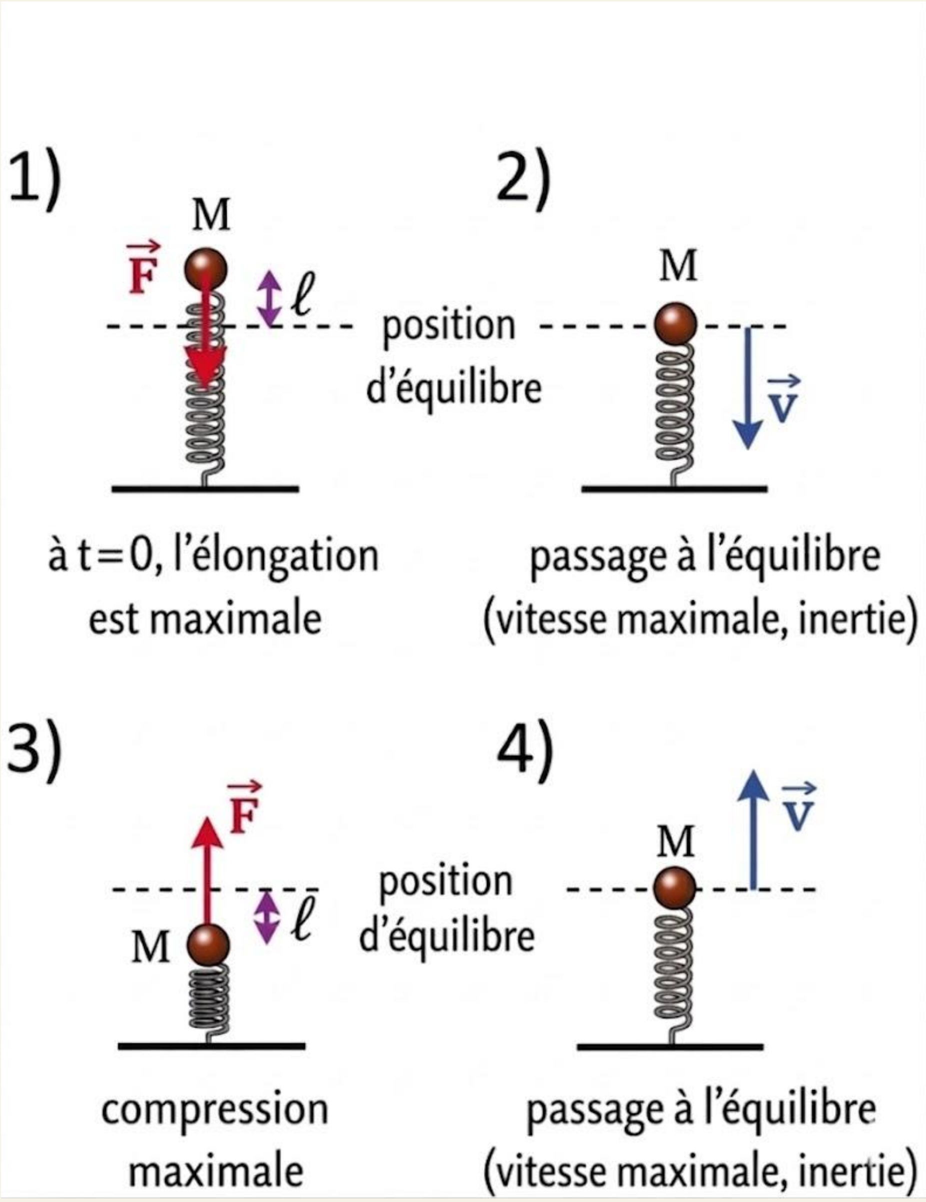


Figure 1

We can already anticipate, without writing down the equations or attempting to solve them, the trajectory followed by the small mass over time. It will have the following form:

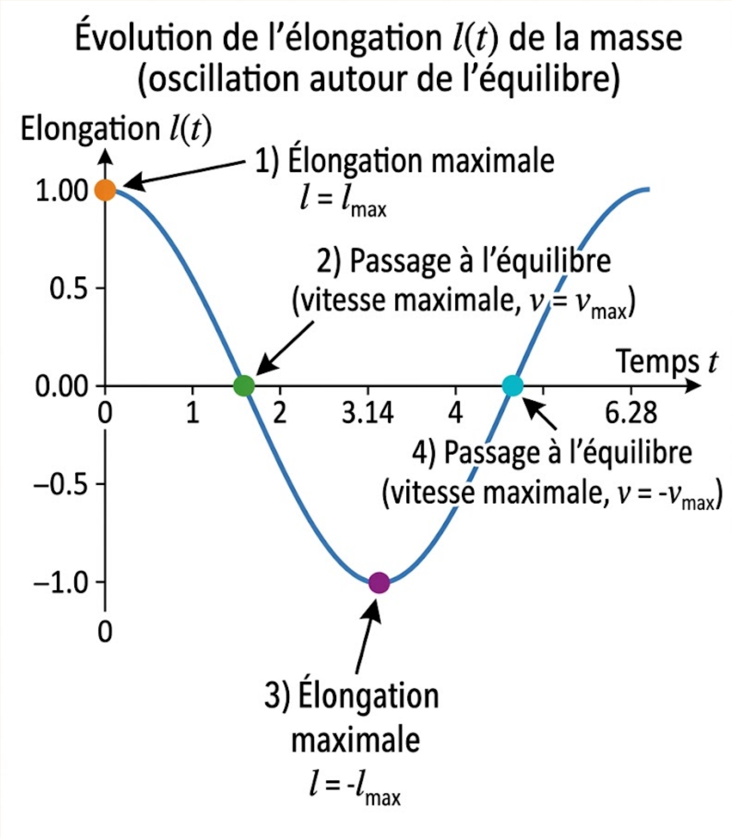


Figure 2

Now that we have an intuitive grasp of the spring's dynamics, we shall formulate them mathematically. We shall proceed slowly, using the opportunity to review a few basic ideas from physics and mathematics. One feature of the model is fundamental: the restoring force is proportional to the displacement ℓ. Recalling Newton's second law, which relates the force acting on an object to its acceleration, we obtain:

$$(1)\quad \overrightarrow{F} = -r\,\overrightarrow{l} = m\,\overrightarrow{a}$$

Since we shall assume that the motion takes place exclusively in the vertical direction, we can temporarily ignore the vector character of force and acceleration. Acceleration is the second derivative of position with respect to time, so the dynamics of the spring are governed by:

$$(2)\quad m\frac{d^2l(t)}{dt^2} = -rl(t)$$

Or equivalently:

$$(3)\quad \frac{d^2l(t)}{dt^2} = -\frac{r}{m}l(t)$$

This is a differential equation, but there is no reason to be intimidated by the term. Its physical meaning is quite intuitive: the second derivative of the mass's trajectory - the rate at which its velocity changes - is proportional to the spring's displacement. Figure 2 makes this immediately visible. At point 1, the displacement is maximal and upward, while the second derivative is maximal and downward. At point 2, both the displacement and the second derivative vanish. At point 3, the displacement is maximal and downward, while the second derivative is maximal and upward. Finally, at point 4, both vanish again.

Mathematically, trajectories with this property can be represented by trigonometric functions such as sine and cosine.

In our case the motion begins at maximum displacement, at t = 0. The solution may therefore be written as a cosine; this choice reflects the initial conditions. We can write the solution as:

$$(4)\ l(t) = l_0 \cos(\omega t)$$

Here ℓ_0 is the spring's initial displacement and, in this case, its maximum displacement. This quantity has another name that will be essential later: the amplitude. The quantity ω measures how rapidly the spring completes an oscillation cycle. It is called the angular frequency.

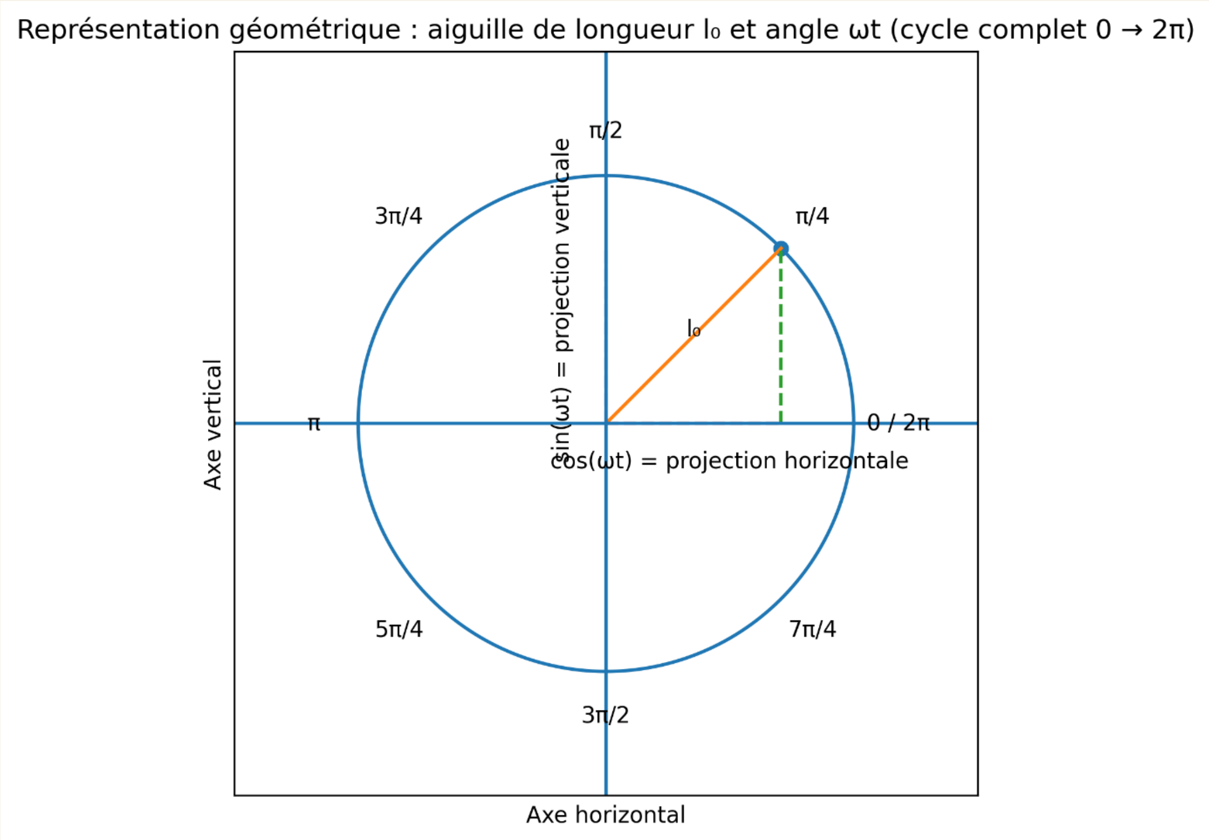


Figure 3: the displacement $\ell(t)$ can be represented as the projection onto the horizontal axis of a pointer of length ℓ_0 rotating anticlockwise with angular frequency ω. The pointer's angle is then ωt, and its horizontal projection is naturally $\ell(t) = \ell_0 \cos(\omega t)$.

We can now relate this representation of the displacement directly to the curve in Figure 2. This will connect the angular frequency to two other fundamental quantities not yet introduced: the period and the frequency.

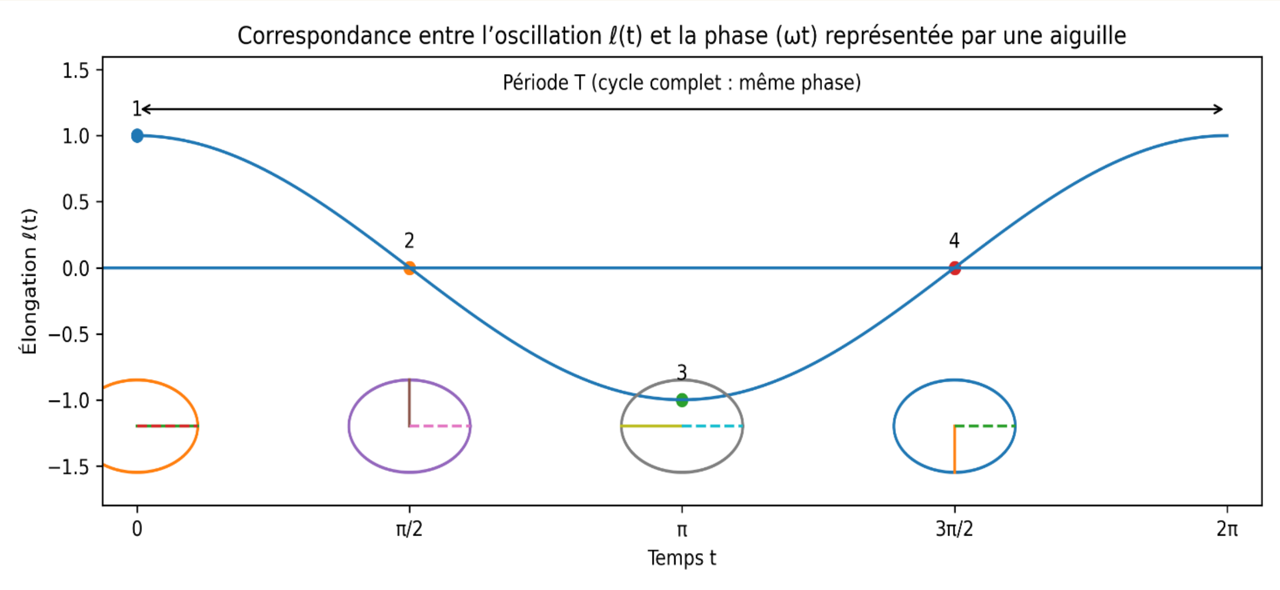


Figure 4: the period T is the time separating two instants at which the spring returns to the same phase of its motion. It is therefore the time required for the 'little clock' to complete one full turn. A complete cycle corresponds to a phase change of 2π. Thus, when the elapsed time equals one period T, the angle has increased by 2π, giving the fundamental relation ωT = 2π.

The diagram shows that when t = T the pointer has completed one revolution. The angle - also called the phase and denoted ϕ - is then equal to 2π. We therefore have a correspondence between T and 2π. By definition, $\phi = \omega t$, up to an additive constant that fixes the arbitrary origin of phase and that we have chosen here to be zero. Hence:

$$(5) \quad \omega T = 2\pi$$

Therefore:

Another important quantity is the frequency ν, defined as the number of cycles completed by the spring per second. It is therefore given by:

$$(7) \quad \nu = \frac{1}{T}$$

We have found the general form of the solution, but we have not yet determined ω or ℓ_0. The value of ℓ_0 depends only on our choice: it is we who decide how far to pull the spring initially and therefore how much energy to supply. The angular frequency, by contrast, is fixed by the equation of motion. To determine it, we differentiate the solution twice with respect to time and substitute the result into equation (3).

$$(8) \quad -\omega^2 l_0 \cos(\omega t) = -\frac{r}{m} l_0 \cos(\omega t)$$

[Reminder: d(cos f(x))/dx = -f'(x) sin f(x), and d(sin f(x))/dx = f'(x) cos f(x).]

The common factors cancel on both sides, leaving:

$$(9) \quad \omega^2 = \frac{r}{m}$$

or equivalently (a very important relation):

$$(10) \quad \omega = \pm\sqrt{\frac{r}{m}}$$

This relation captures much of the physics of the harmonic oscillator. It tells us several things:

- The signs + and - determine the direction in which the spring traverses its cycle. Physically, beginning from top to bottom rather than from bottom to top merely introduces a phase shift - a change in the initial conditions. The choice of sign therefore carries no important physical content here.

- More importantly, ω does not depend on ℓ_0. This is a remarkable property of the harmonic oscillator: its angular frequency is independent of the system's initial state and depends only on the system's constant parameters, r and m. The stiffer the spring, the higher its natural angular frequency; the larger the mass, the lower it is. Because this angular frequency is a constant of the free motion, we give it a special notation, ω_0. The subscript 0 indicates that it is the spring's natural angular frequency: the frequency at which it oscillates when left free, with no external force disturbing its motion.

I.3 Energy of a Free Harmonic Oscillator

We still need to express one important quantity associated with the moving mass: its energy. In classical mechanics, the energy of a point particle is the sum of its kinetic and potential energies. For a harmonic oscillator it is:

$$(11) \qquad E = E_c + E_p = \frac{1}{2}m(\frac{dl}{dt})^2 + \frac{1}{2}rl^2$$

This expression is stated here without proof, but it is not arbitrary. Starting from the work done by the force, one can show that this quantity is conserved (see the work-energy theorem). It shows that the energy stored in the system depends on the square of the velocity and on the square of the spring's displacement.

Substituting solution (4) into equation (11), a short calculation gives the energy in terms of the amplitude ℓ_0:

or

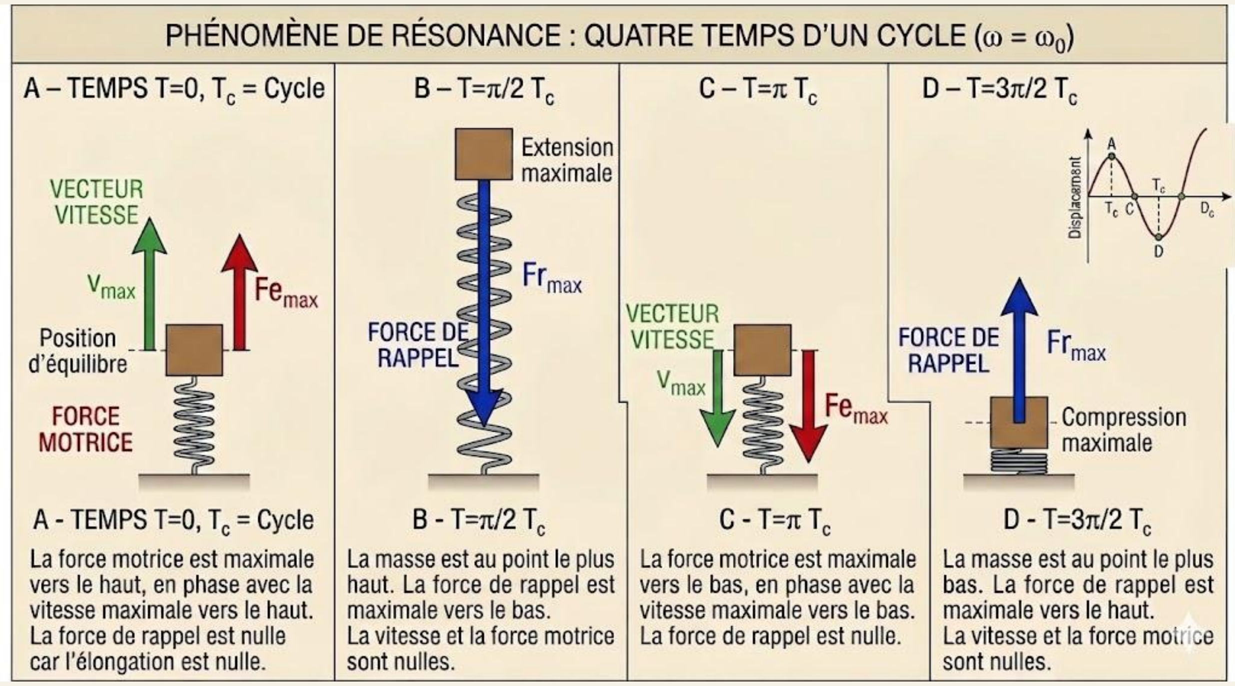
We obtain an important result: the energy is proportional to the square of the amplitude. On reflection this is quite natural: at maximum displacement the kinetic energy vanishes because the velocity is zero, and the total energy is simply the spring's potential energy.

I.4 Linearity

Equation (3) is linear. This means that if we have several solutions of the equation, any linear combination of those solutions is also a solution. More concretely, if several sine and cosine functions of the same angular frequency are solutions, then any weighted sum of them is a solution as well. Linearity is a central property throughout physics. For a single harmonic oscillator it does not generate genuinely new dynamics, because every such combination can be rewritten as one oscillation with a different amplitude and phase. There is only one allowed natural angular frequency. We introduce linearity here because it will become indispensable when we turn to coupled oscillators and fields.

I.5 Resonance

One of the most striking phenomena in physics is resonance. It occurs when a physical system with one or more natural frequencies is driven by a periodic disturbance matching one of those frequencies. For our simple spring, resonance therefore occurs when the driving angular frequency approaches the spring's natural angular frequency. Qualitatively, the mechanism is straightforward. Resonance occurs when the driving force and the spring's velocity remain synchronised and point in the same direction. The rate at which a force transfers energy to an object - the power - is the scalar product F·v. Synchronisation therefore keeps the transferred power positive and close to its maximum. From one cycle to the next, more energy is injected into the spring: its maximum kinetic energy increases, its displacement grows, and so does its amplitude.



Mathematically, if an external periodic force of angular frequency ω drives the spring, the equation of motion becomes:

$$(13) \quad \frac{d^2 l(t)}{dt^2} = -{\omega_0}^2 l(t) + \frac{F_0}{m}\cos(\omega t)$$

When $\omega \neq \omega_0$, the motion is the sum of two oscillations, one at ω and the other at ω_0. Let us isolate the oscillation at the imposed angular frequency ω. We try a cosine solution, called the spring's forced response. In a weakly damped real oscillator, the system would eventually converge towards this regime. Here, without damping, the free oscillation at ω_0 never disappears, but it is the forced response that carries the resonance. We write it as:

$$(14) \quad l(t) = A\cos(\omega t)$$

Substituting this solution into the equation gives:

$$(15) \qquad -\omega^2 A\cos(\omega t) = -{\omega_0}^2 A\cos(\omega t) + \frac{F_0}{m}\cos(\omega t)$$

Hence:

$$(16) \quad A = \frac{F_0}{m({\omega_0}^2 - \omega^2)}$$

Something remarkable appears: the amplitude of the forced response tends towards infinity as ω approaches ω_0. We must not, however, make a tempting mistake. Equation (16) describes no physical state with an actually infinite amplitude. It combines two limiting processes: ω approaches ω_0 and the elapsed time tends towards infinity. Even an ideal harmonic oscillator would require an infinite time to reach an infinite amplitude. At exact resonance ($\omega = \omega_0$), moreover, the forced response is no longer a simple cosine with constant amplitude: its amplitude grows linearly with time, as $t \cdot \sin(\omega_0 t)$. Finally, the sign of A reverses as ω passes through ω_0. For $\omega > \omega_0$, A is negative, indicating that the oscillation is in antiphase with the driving force. The essential point is that the energy transferred by an external force to an oscillating system becomes enormously greater when the driving angular frequency approaches the system's natural angular frequency.

II. A Chain of Coupled Harmonic Oscillators

II.1 From a Spring to a Wave

The physics of a single isolated harmonic oscillator is relatively limited and can be explored quite quickly. We shall now study a much more interesting system, one that will later help us understand field physics in depth. The system nevertheless remains very simple and could almost be built at home without too much difficulty. You need a series of identical springs fixed side by side to a board at equal intervals. Attach an identical small mass to the free end of each spring. The final and decisive step is to connect every mass to its two neighbours with elastic bands.

To make our lives easier, we shall introduce several idealisations. The restoring force will be exactly proportional to displacement; the springs will also be perfectly rigid laterally, so that once they are fixed to the board the masses can move only in the vertical direction. The elastic bands will likewise be ideal, with restoring forces exactly proportional to their extension. None of these materials will produce sound or heat, and the experiments will take place in weightlessness. These liberties allow us to work with the simplest possible equations.

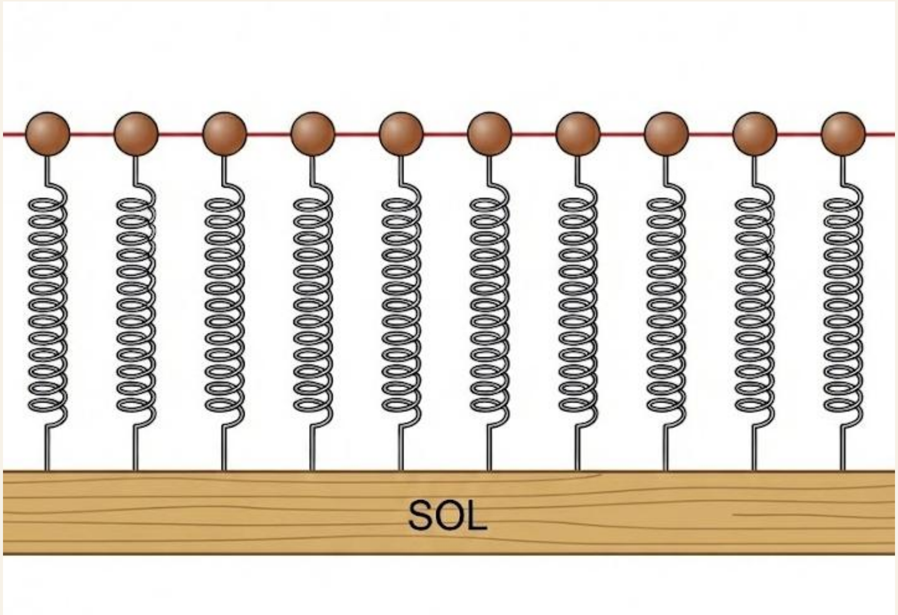


Figure 6: A chain of harmonic oscillators coupled by elastic bands

What happens if we pull one of the small masses upwards and then release it? Qualitatively, we should expect the following.

In the initial configuration:

- The displaced mass is higher than its neighbours, so the elastic bands exert an upward force on them.
- Its immediate neighbours on the left and right are therefore pulled upwards as well, though not as far as the mass that was directly displaced.
- The neighbour on the right in turn exerts an upward force, through the elastic band, on its own neighbour to the right. The same happens on the left.
- And so on…

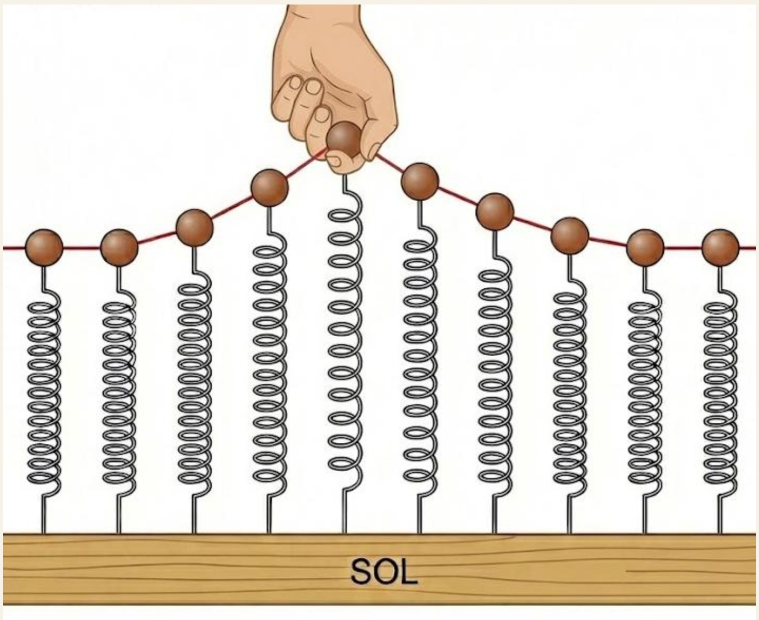


Figure 7: One mass is displaced.

Upon release:

- A fairly complex evolution follows. Each mass is acted upon by three forces: (1) the restoring force of its own spring, which depends on the spring's extension or compression; (2) the force exerted by the elastic band on its right, directed upwards if the neighbour on the right is higher and downwards if it is lower; and (3) the corresponding force exerted by the elastic band on its left.
- The resultant force also depends on the respective stiffnesses of the elastic bands and the springs.
- Initially, however, every mass experiences a downward force. The largest force acts on the displaced mass, because it experiences both the strongest restoring force from its own spring and the downward pull transmitted through the two elastic bands connected to its neighbours.

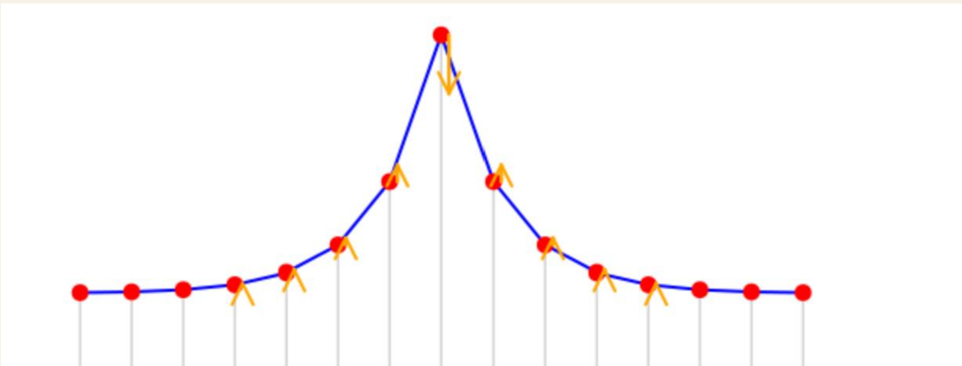


Figure 8: t = 0. The yellow vectors represent the resultant elastic-band force on each mass. The restoring forces of the vertical springs are not shown.

- All the masses nevertheless begin to oscillate. The initially displaced mass has the greatest amplitude, and the amplitude becomes smaller with increasing distance from it.
- Because the masses do not experience the same forces at every instant, however, they gradually fall out of phase. The central mass progressively loses amplitude because the elastic forces generated by its neighbours are not synchronised with its velocity.

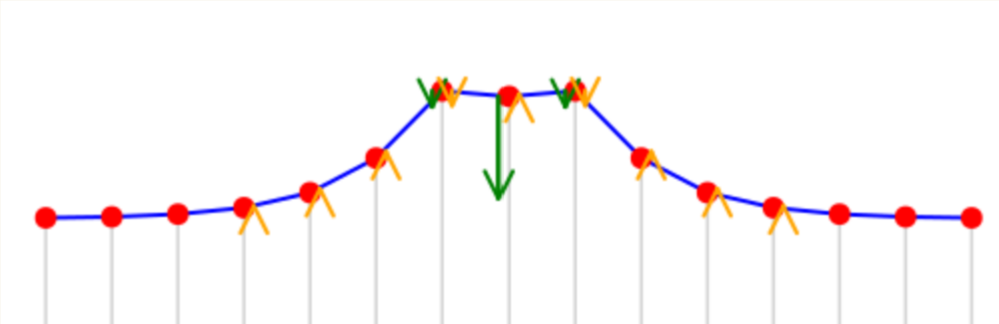


Figure 9: t = 1. The green vectors represent the velocities of the small masses. For the central oscillator, the velocity and the elastic force point in opposite directions.

- Its immediate neighbours, by contrast, temporarily gain amplitude because the force exerted by the central mass is momentarily aligned with their velocity: a temporary resonance condition.

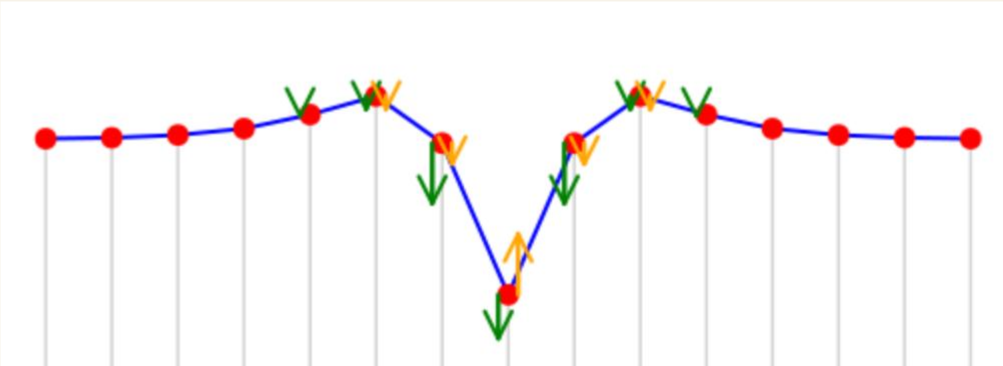


Figure 10: t = 2. The elastic force opposes the velocity of the central oscillator, but is aligned with the velocities of the neighbouring oscillators.

- The neighbouring oscillators then lose amplitude in turn when the elastic forces and their velocities cease to be synchronised. Their own neighbours subsequently benefit from the transferred energy.

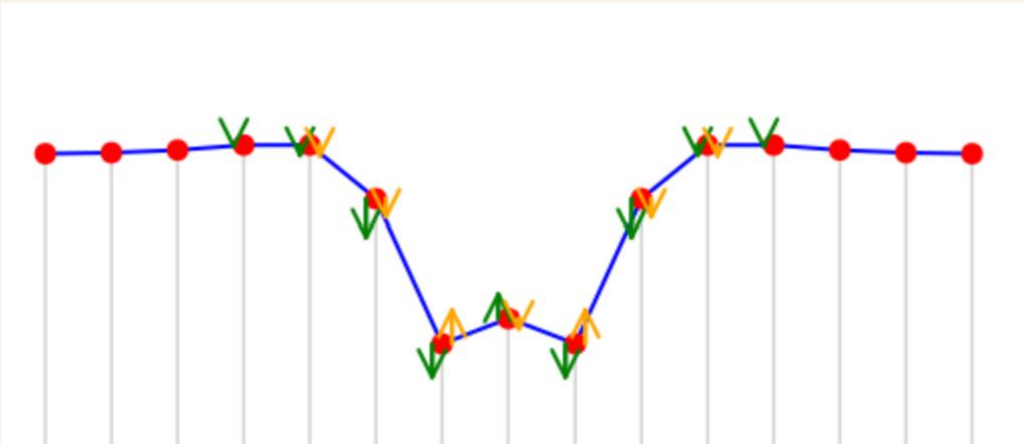


Figure 11: t = 3.

- Through this continuously evolving pattern of phase differences between spring velocities and elastic forces, the conditions for resonant energy transfer shift step by step through space.

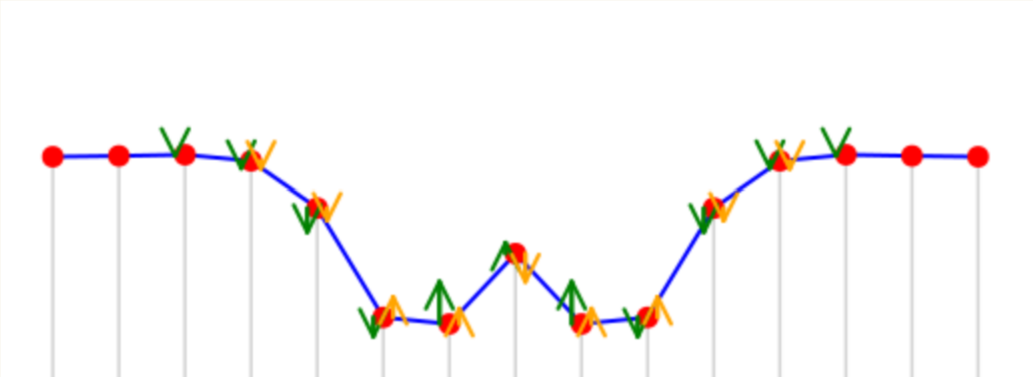


Figure 12: t = 4.

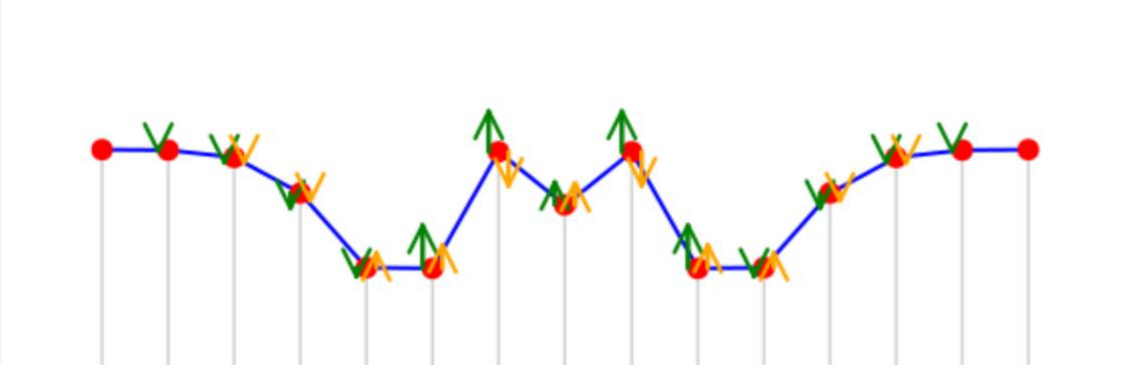


Figure 13: t = 5.

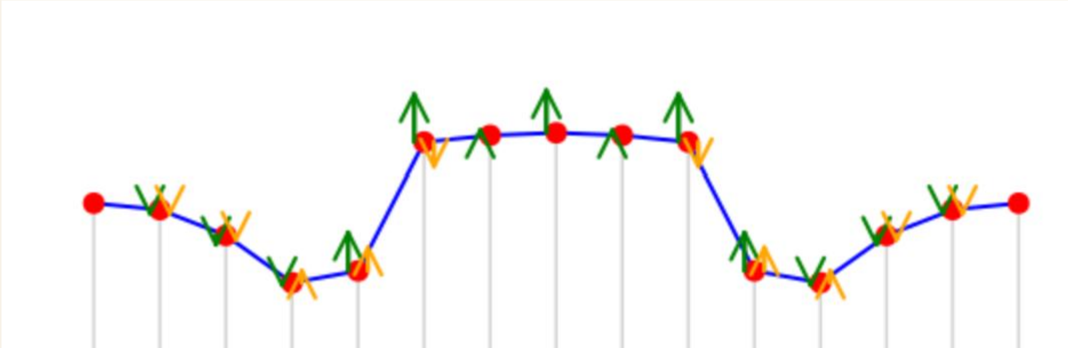


Figure 14: t = 6.

- From an energetic point of view, it is as though the energy initially concentrated at the centre begins to propagate both to the right and to the left. More precisely, each oscillator gains energy when the elastic forces and its velocity point in the same direction, and loses energy when they point in opposite directions. This is the precise mechanism by which energy spreads through the network of springs. Energy is not destroyed; it is progressively transferred from one oscillator to its neighbours by the coupling forces of the elastic bands.

(see Interactive Animation 1)

This propagation of energy through space without any overall transport of matter - every spring remains in place - is what we call a wave. More precisely, a wave is the propagation of a disturbance through a medium or through space; here, the initial disturbance is the displacement of the central oscillator. The concept of a wave is one of the most important in physics. It describes a collective phenomenon: in our model, it depends on the organised motion of several springs connected by elastic bands. This organisation allows us gradually to set aside the underlying springs and treat the wave itself as the central object. Although it emerges from the collective dynamics of the oscillators, the wave begins, in a sense, to acquire a life of its own.

A familiar example is a surface wave on water, which consists of local disturbances in the height of the liquid surface. The water molecules at the surface do not travel with the wave. The wave thus appears to become partly independent of the water that constitutes it, although at the fundamental level it is nothing more than the organised collective motion of water molecules. Two complementary languages therefore emerge. The more fundamental description speaks in terms of water molecules; the emergent description speaks in terms of waves.

The spring model we have just examined is, in fact, not a bad first model of a wave on the surface of water. The small masses would represent surface molecules, the elastic bands the intermolecular bonds between them, and the vertical springs the interactions connecting surface molecules to molecules slightly deeper in the liquid. This would of course be a very crude model, not least because molecules in a liquid possess many more degrees of freedom.

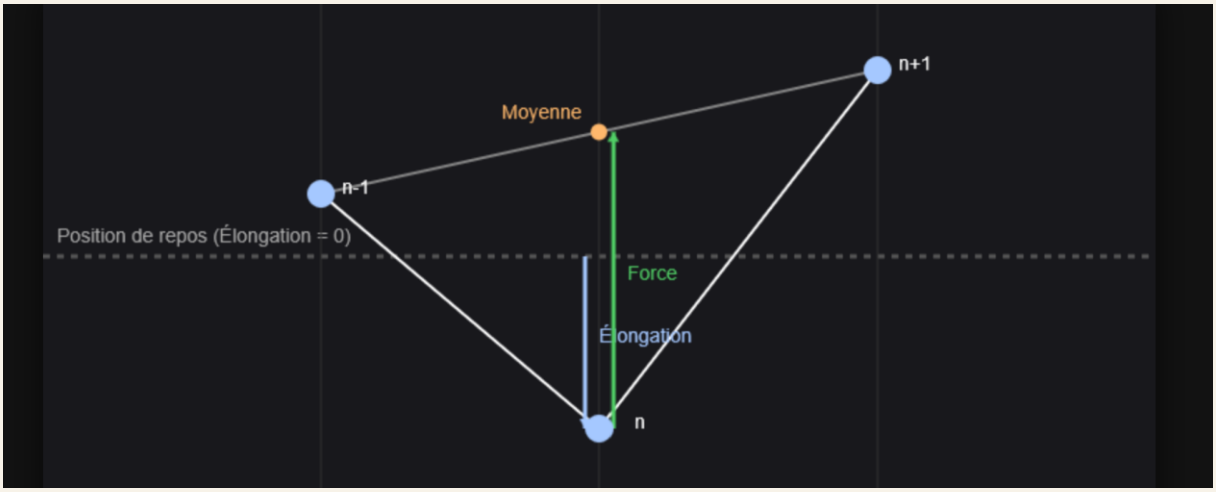
II.2 Local Curvature and the Force Exerted by the Elastic Couplings

What follows is crucial. Stay with me: it is not difficult. We shall now arrange our chain of springs in a circle. This means that the rightmost oscillator is connected to the leftmost one, so that a disturbance leaving the chain on one side re-enters it from the other. We are looking for configurations that can persist indefinitely without changing their spatial form.

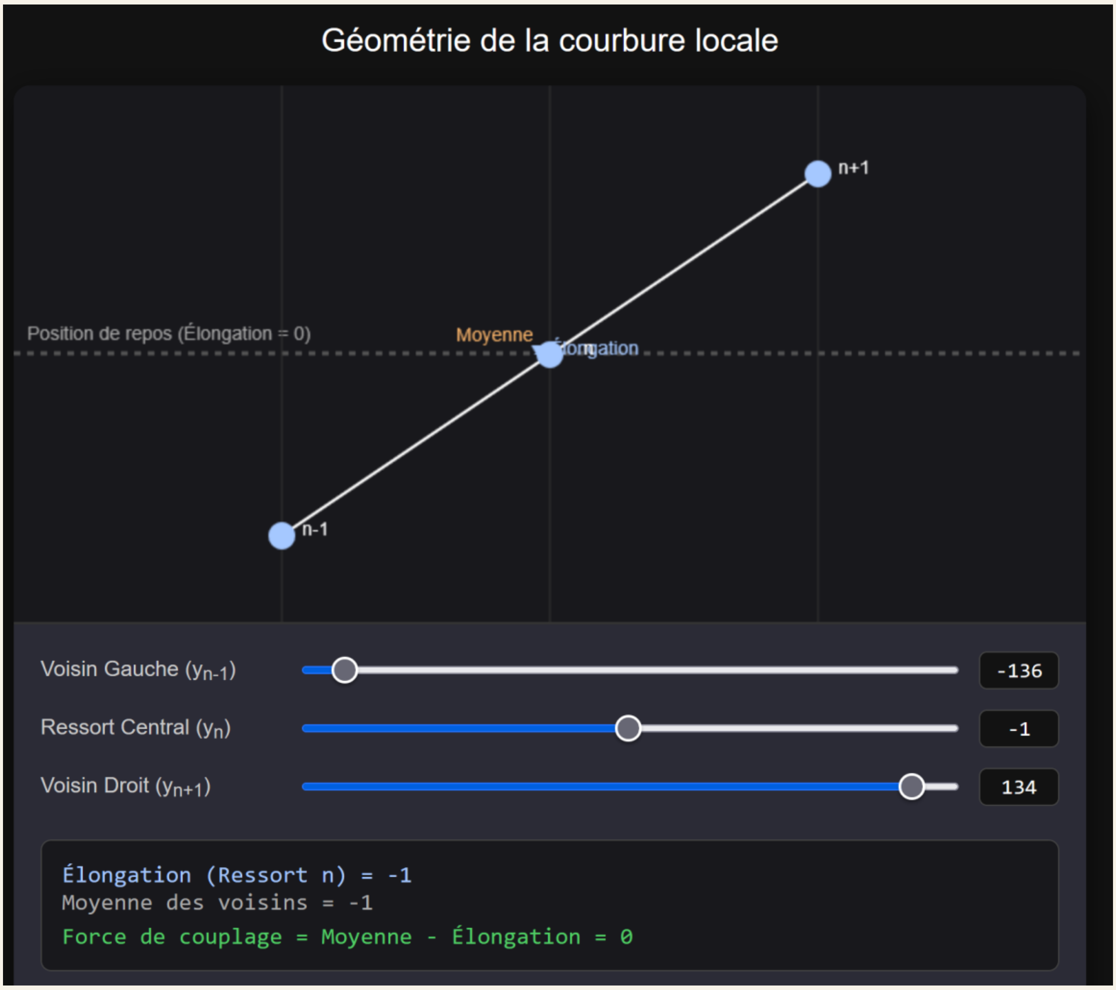
One solution is very easy to find: all the oscillators may be perfectly synchronised, with the same amplitude and the same phase. In that case, every elastic coupling remains unstretched, so the only force acting on each mass is the restoring force of its own spring. All the masses then move together as copies of the same harmonic oscillator.

Are there other solutions? Yes. The force exerted by the elastic couplings must simply 'simulate' an additional restoring force. For each oscillator, the resultant elastic force must therefore be proportional to the negative of that oscillator's displacement.

How can the elastic couplings produce such a force on every oscillator? Notice that the resultant elastic force on mass n depends on the difference between its position and the average position of its two neighbours, n-1 and n+1.



That difference is exactly a discrete expression of local curvature. When the three masses are aligned, the difference between the position of n and the average position of its neighbours is zero: the three points have no local curvature.



In that case, the forces exerted by the two elastic couplings balance one another and the resultant force is zero, as the figure above shows. Local curvature is defined by the departure of the central point from the straight line determined by its neighbours. The larger this departure, the larger the resultant elastic force.

(see Interactive Animation 2)

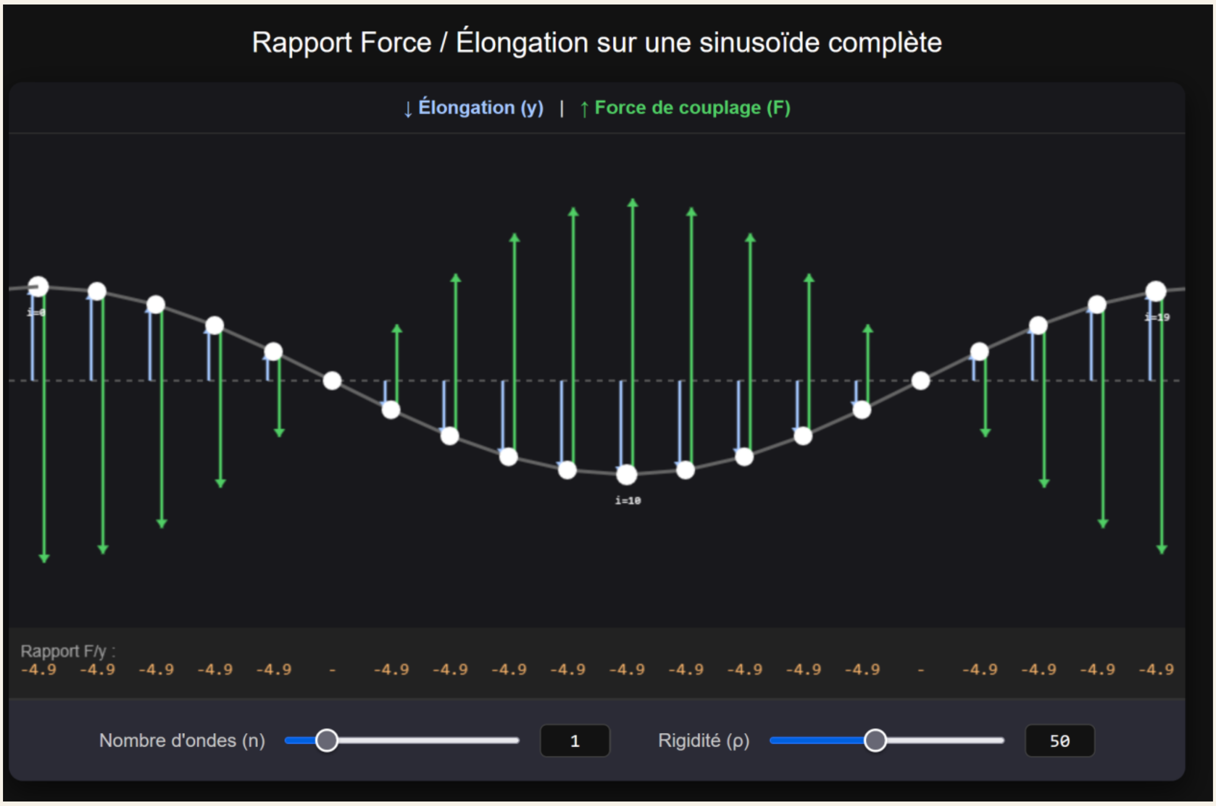
Let us summarise what has just been established:

1) We are looking for a solution in which, for every oscillator, displacement is proportional to the negative of the resultant elastic force.
2) That force is itself proportional to the local curvature of the configuration.

Conclusion: we are looking for a solution in which displacement is proportional to the negative of its own local curvature.

Recalling that the local curvature of a function is represented by its second derivative, there is one family of functions that satisfies this requirement perfectly: sine and cosine functions. Their second derivative is proportional to the original function, with the opposite sign.

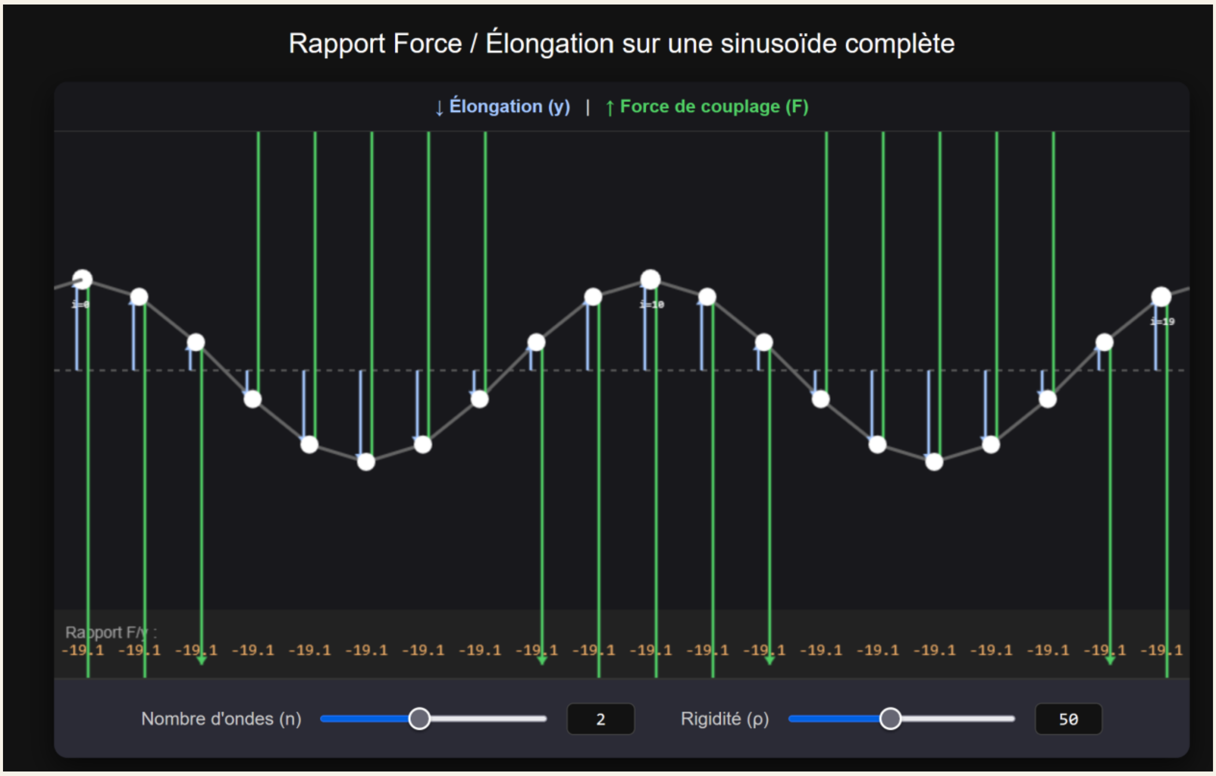
To verify this, the diagram below displays the ratio between force and displacement over one complete spatial cycle:



The yellow values at the bottom show that this ratio is the same for every oscillator.

We can now change the length of the cycle and check whether the same property is preserved:

(see Interactive Animation 3)



Here the distance between two successive peaks is smaller, and the magnitude of the ratio has increased (here, -19.1). This is exactly what we should expect, because the curvature is now greater. The essential point, however, is that the ratio remains constant across all the oscillators. A sinusoidal configuration therefore makes the elastic coupling force behave like an additional restoring force.

At this stage, however, we have not yet shown that if these conditions are satisfied initially, they will remain satisfied at all later times. That will depend on the initial velocity distribution, a point to which we shall return.

II.3 Wavelength, Quantisation, and Modes

A second difficulty now arises: we stated that the network is arranged in a circle. In that case, the required conditions must remain satisfied when we pass from the rightmost oscillator (let us call it Marine) to the leftmost oscillator (let us call it Jean-Luc). This is possible only if an exact integer number of spatial cycles fits around the circle. One spatial cycle is what we call **a wavelength**, and it is denoted by the Greek letter λ. For a circle of circumference L, the spatial condition for a normal mode is: $L = n\lambda_n$.

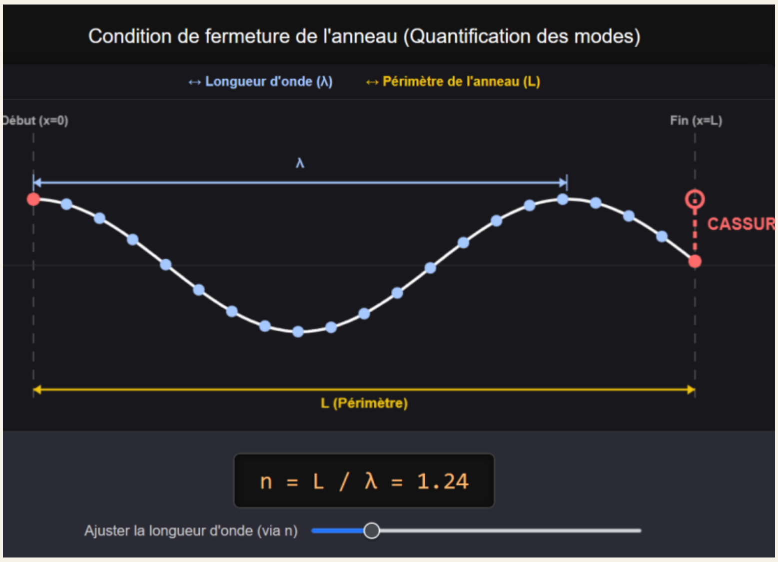


Figure 14: the periodic condition is not satisfied here. The ratio is not an integer, and a discontinuity appears at the junction.

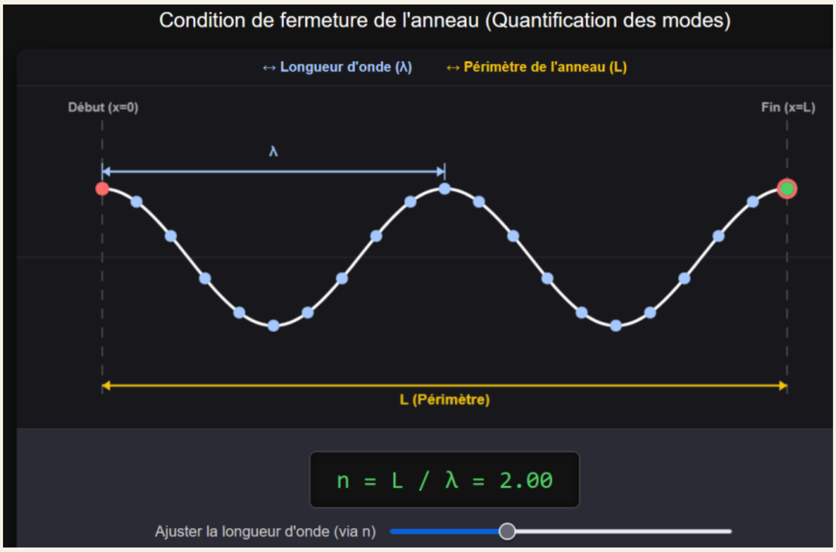


Figure 15: here the condition is satisfied. n is an integer, and the configuration is continuous at the junction.

(see Interactive Animation 4)

Just as the temporal period is related to angular frequency - the rate of change of phase per unit time - by the relation $\omega = \frac{2\pi}{T}$, called **wavenumber** and denoted by k, through the relation: $k = \frac{2\pi}{\lambda}$. The wavenumber therefore represents the rate of change of phase per unit length.

Waves of this kind, corresponding to permanent regimes, are called monochromatic waves λ_n because they contain only one wavelength. Each monochromatic wave satisfying the condition $L = n\lambda_n$ is called **a mode.** The lowest-energy mode, corresponding to the lowest frequency, is mode 0, also called **the fundamental mode (in the sense of the lowest-frequency mode; not to be confused with the first harmonic of a string or pipe)**, for which the wavelength is infinite. It corresponds to the configuration in which all the oscillators are at the same height. In that case, $\varrho = 0$ and $\omega = \omega_0$.

Figure 16: fundamental mode n = 0. The elastic coupling force vanishes, so the effective restoring force on each mass, and therefore the frequency, is minimal.

The larger n is, the smaller λ_n becomes, and the larger ϱ becomes. A small λ_n favours larger departures from a straight line in the spatial distribution of the masses, and therefore larger resultant forces from the elastic couplings. The effective restoring force on each oscillator thus increases, and so does its frequency. The fact that **shorter wavelengths** correspond to **higher frequencies** is a general property of waves.

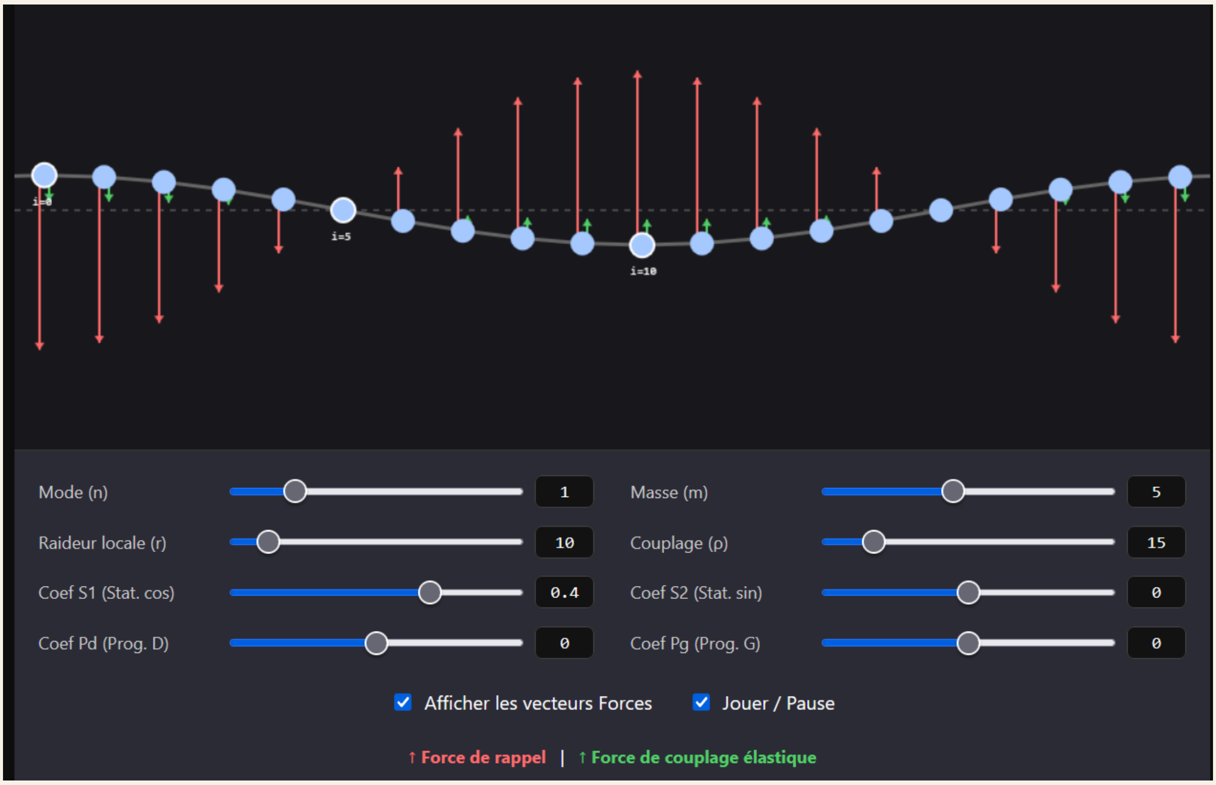


Figure 17: mode 1. The elastic coupling force is present but remains weak, so the frequency increases only slightly.

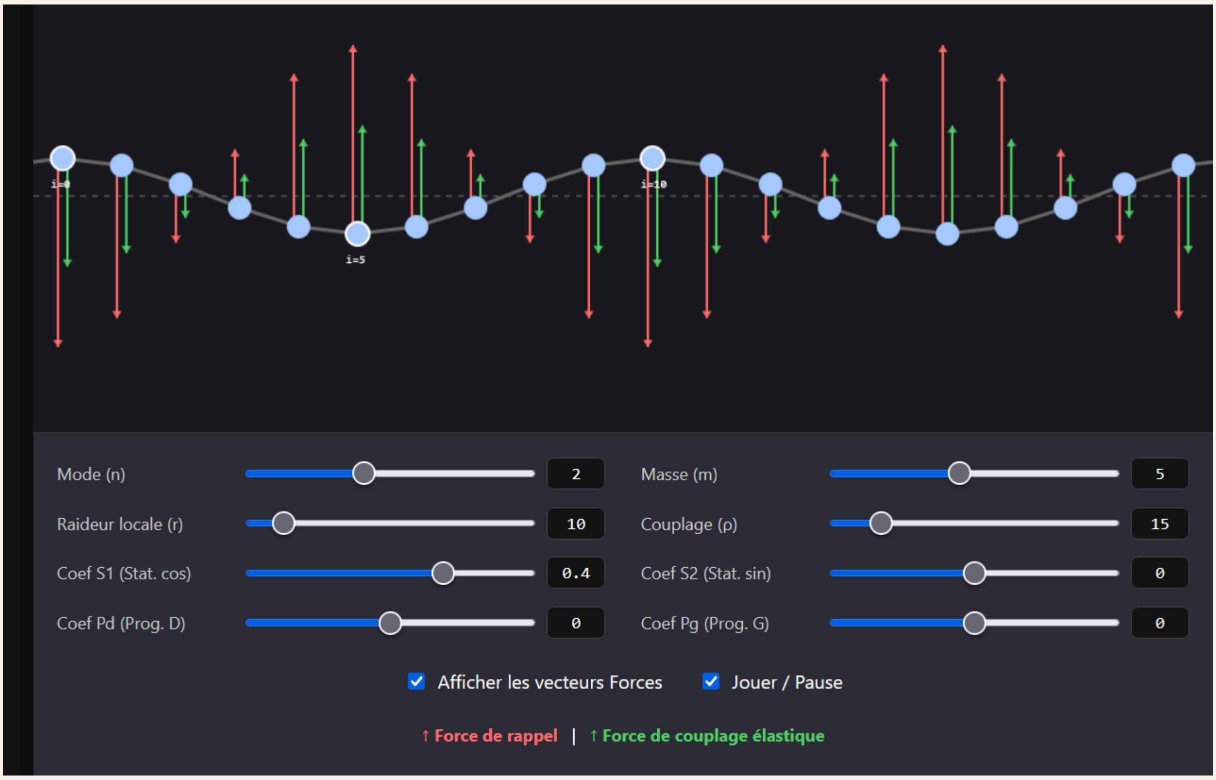


Figure 18: mode 2

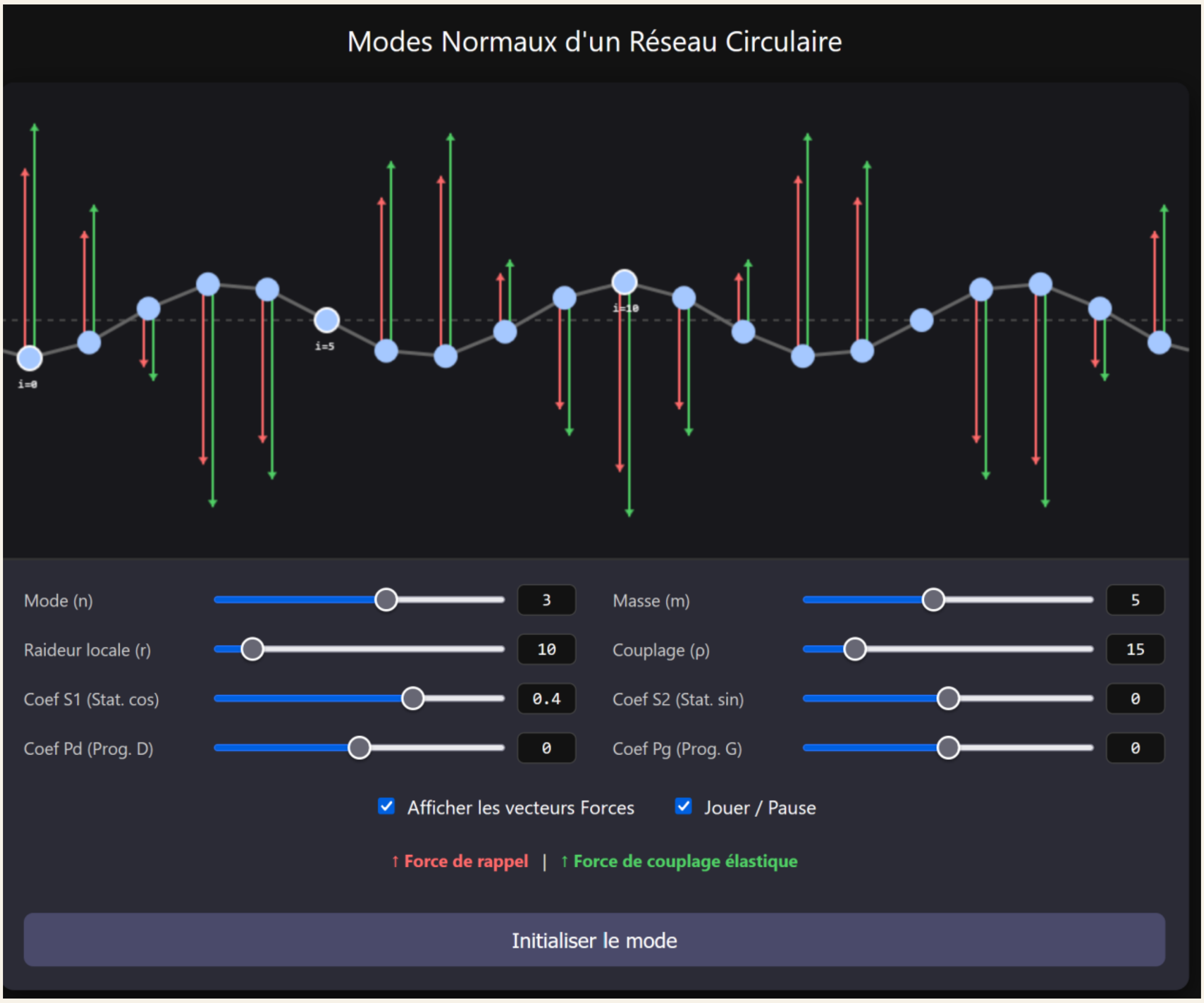


Figure 19: mode 3

Modes are crucial because they are the basic ingredients of every wave. We can even venture an analogy: just as a material object is composed of different kinds of atoms in different proportions, a wave is composed of different modes in different proportions. The analogy is not exact, but it captures the central idea: modes form the elementary building blocks from which more complex waveforms can be constructed.

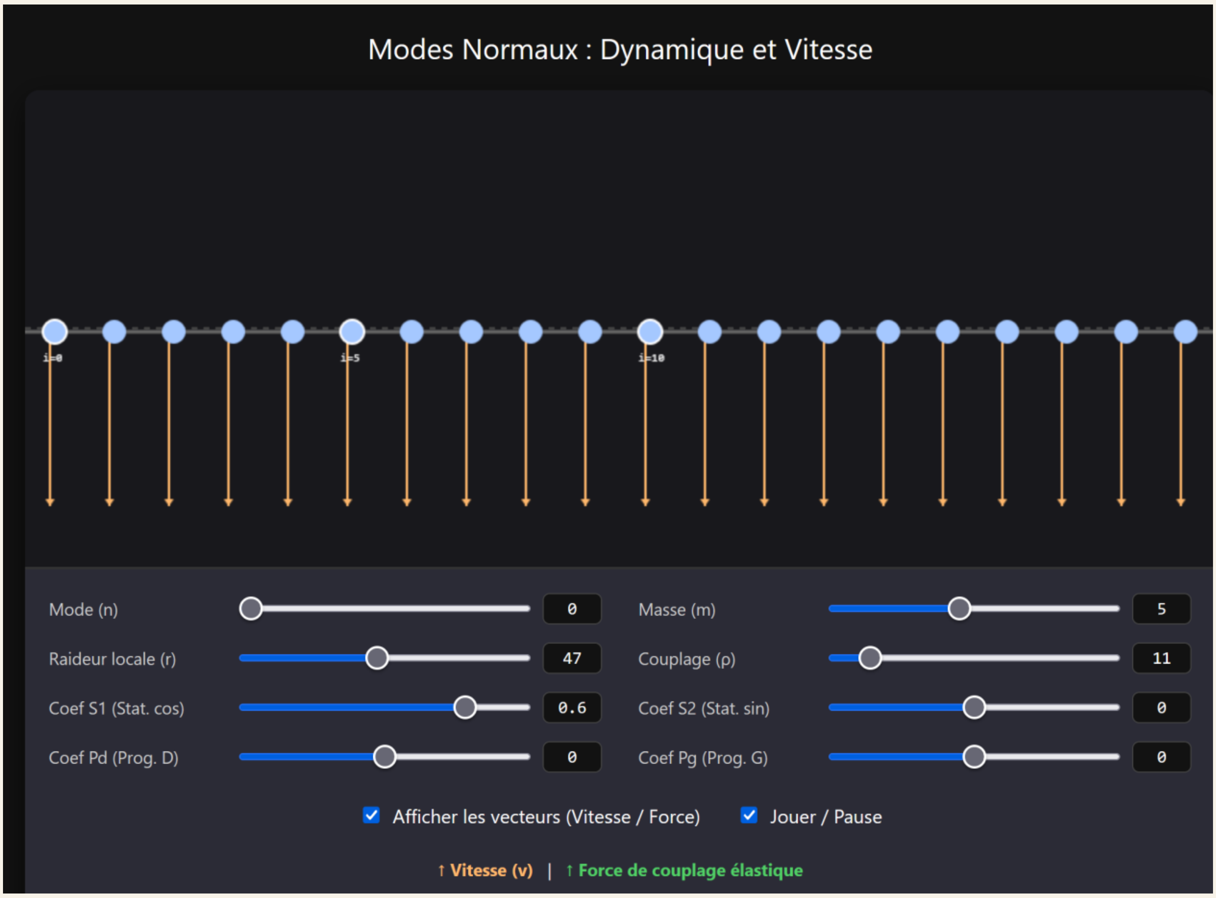
II.4 Mode Degeneracy

We shall continue our survey of wave properties. We have identified the conditions under which a wave can persist indefinitely in a closed medium of length L:

- Each spring must obey a harmonic-oscillator equation at all times. This is possible only if the force exerted by the elastic couplings 'imitates' a restoring force: the spatial arrangement of the springs must therefore be sinusoidal.
- The sinusoidal function must have a spatial period—that is, a wavelength—such that $L = n\lambda_n$, where n is an integer.

However, we have not yet explained why, **once the initial configuration is sinusoidal, it should remain sinusoidal**. This is not automatic. It holds only under certain conditions, which concern the way the initial velocities of the small masses are distributed.

Mode 0:



This is the simplest mode. All the masses have the same height, so the elastic couplings are not stretched. If they are launched with the same velocity (the yellow vectors above), they are guaranteed to remain at the same height at every later time. In other words, there is only one possible velocity distribution: the velocity must be identical everywhere. Because there is only one such condition, **mode 0 is non-degenerate**. The forbidding term simply means that mode 0 is unique in its overall spatial form, although its amplitude and initial velocity may of course vary. The fundamental reason is that, for k = 0, sin(0·x) = 0 identically: the only non-trivial sinusoidal function of infinite wavelength is the constant cos(0·x) = 1. The solution space for mode 0 is therefore one-dimensional.

Mode $n \neq 0$:

When we studied the harmonic oscillator, we did not emphasise one point. We examined the trajectory of the small mass over time—for example, a function of the form cos(ωt)—but did not discuss the form of its velocity over time. Differentiating cos(ωt) with respect to time gives −ω sin(ωt). More generally, the derivative of a sinusoidal function is another sinusoidal function with the same period, but shifted in phase by π/2. Thus, when the displacement reaches its maximum upward value, the velocity is zero; when the displacement is zero, the velocity is maximal and directed downwards; when the displacement reaches its maximum downward value, the velocity is zero again; and so forth.

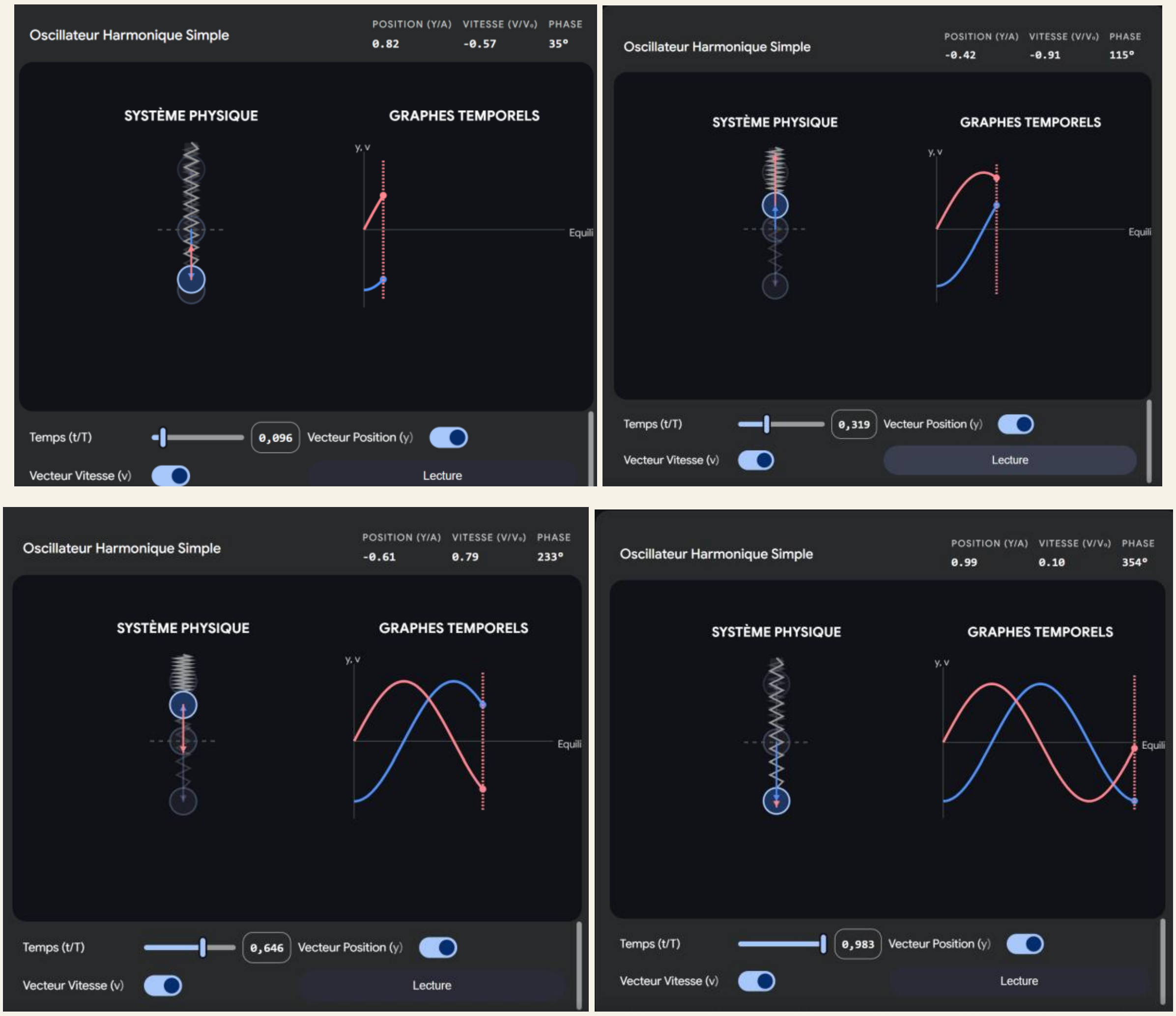


Figure 20: position and velocity vectors over time.

(see Interactive Animation 6)

We shall now generalise this logic—not to temporal evolution, but to the spatial configuration. If the initial spatial configuration of the springs is sinusoidal, and if the spatial distribution of velocities is also sinusoidal with the same spatial period—that is, the same wavelength—then both the spring configuration and the velocity configuration remain sinusoidal.

To see why, recall that if we have a function l(x,t), then at time t+dt we have l(x,t+dt)=l(x,t)+[∂l(x,t)/∂t]dt. But ∂l(x,t)/∂t is, by definition, the spring-velocity function as a function of x. If this velocity is itself sinusoidal and has the same wavelength as l(x,t), then l(x,t+dt) is the sum of two sinusoidal functions with the same wavelength. The sum of two sinusoidal functions with the same wavelength is itself sinusoidal with that same wavelength, as the following diagram shows:

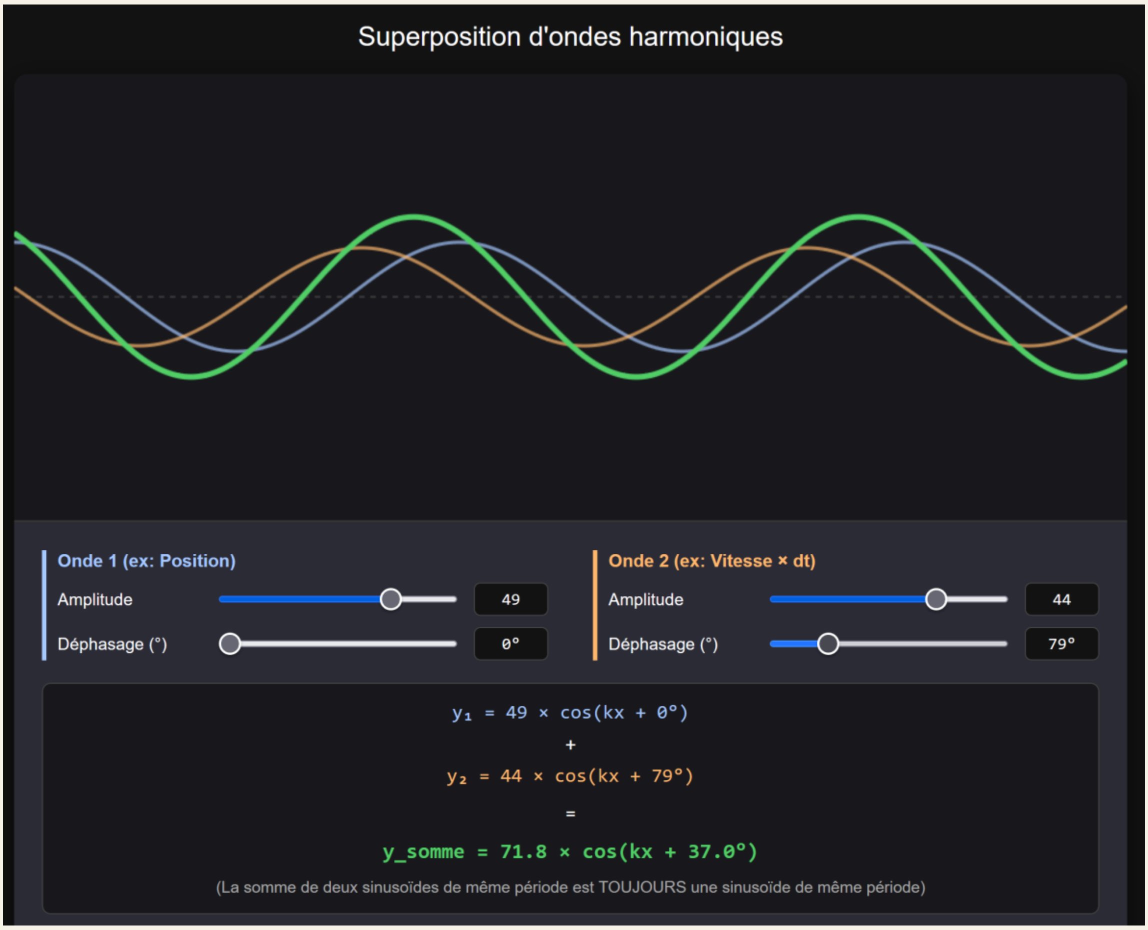


Figure 21: sum of two sinusoidal functions with the same period.

(see Interactive Animation 7)

The evolution of l(x,t) over time is nothing more than an infinite iteration of such infinitesimal sums, each preserving the sinusoidal form and the original wavelength. We must nevertheless verify that v(x,t) also remains sinusoidal at every step; otherwise the argument cannot be iterated. This is where the equation of motion enters. The acceleration of each mass is proportional to the negative of its displacement, as established in Section II.2. Since l(x,t) is sinusoidal, the acceleration is sinusoidal as well. Therefore v(x,t+dt)=v(x,t)+a(x,t)dt is the sum of two sinusoids with the same wavelength and thus remains sinusoidal. The loop is closed.

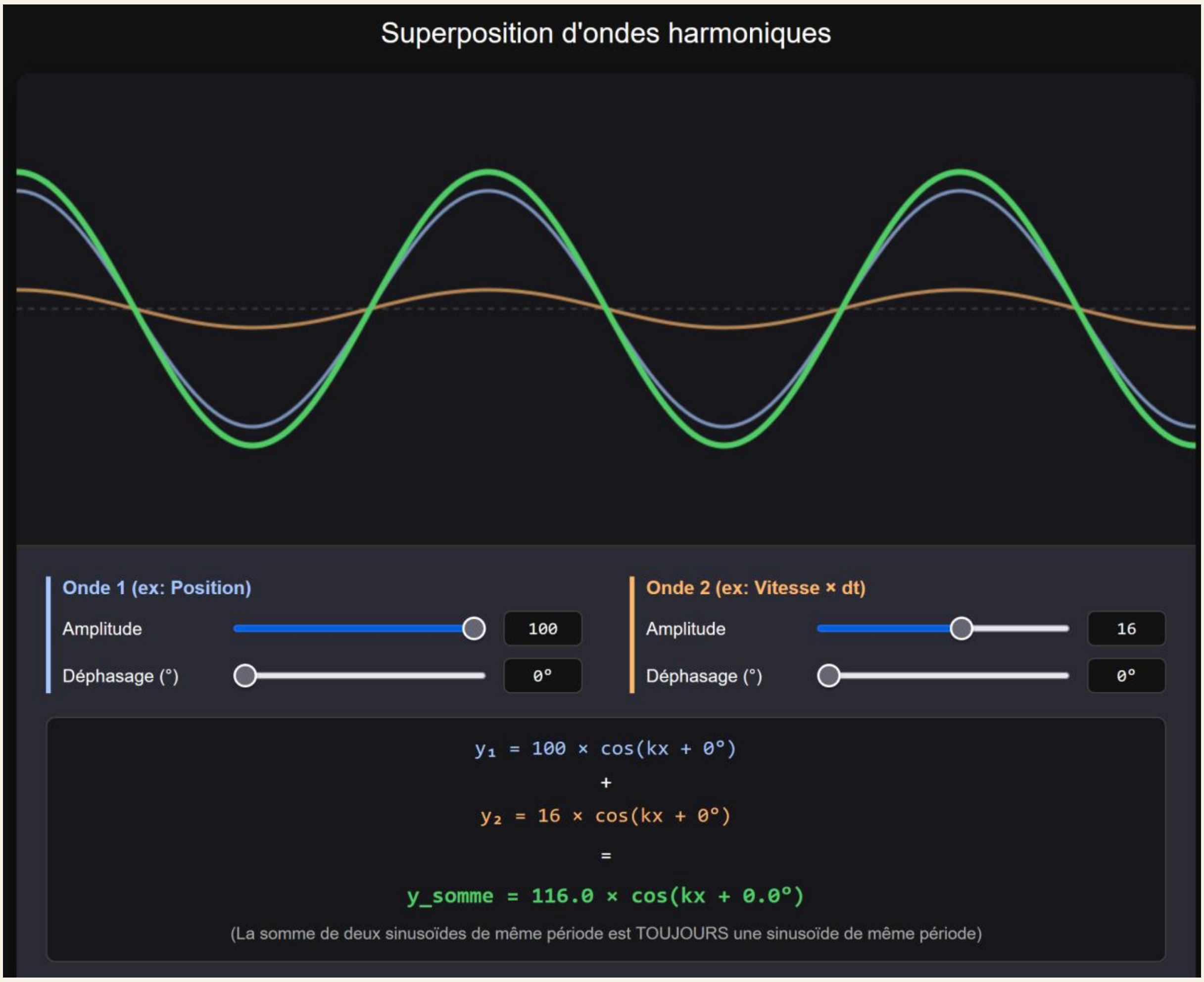
To guarantee this preservation, all we need do is require l(x,t) and v(x,t) to be sinusoidal functions with the same wavelength in the initial state. We are nevertheless free to choose the spatial phase difference between l(x,t) and v(x,t). The temporal phase difference, by contrast, is necessarily π/2. There are therefore several solutions satisfying these conditions. Although infinitely many solutions are possible, they are not all independent. For each n>0, every solution can be constructed as a linear combination of basis solutions. These modes are said to be 'degenerate': for a given n, several independent solutions share the same frequency ω. We shall focus on four special cases. They are the ones most often discussed in the literature, and they stand in particularly revealing relations to one another. These relations will later allow us to uncover something extremely deep—indeed, almost abyssal—about wave modes.

(see Interactive Animation 5)

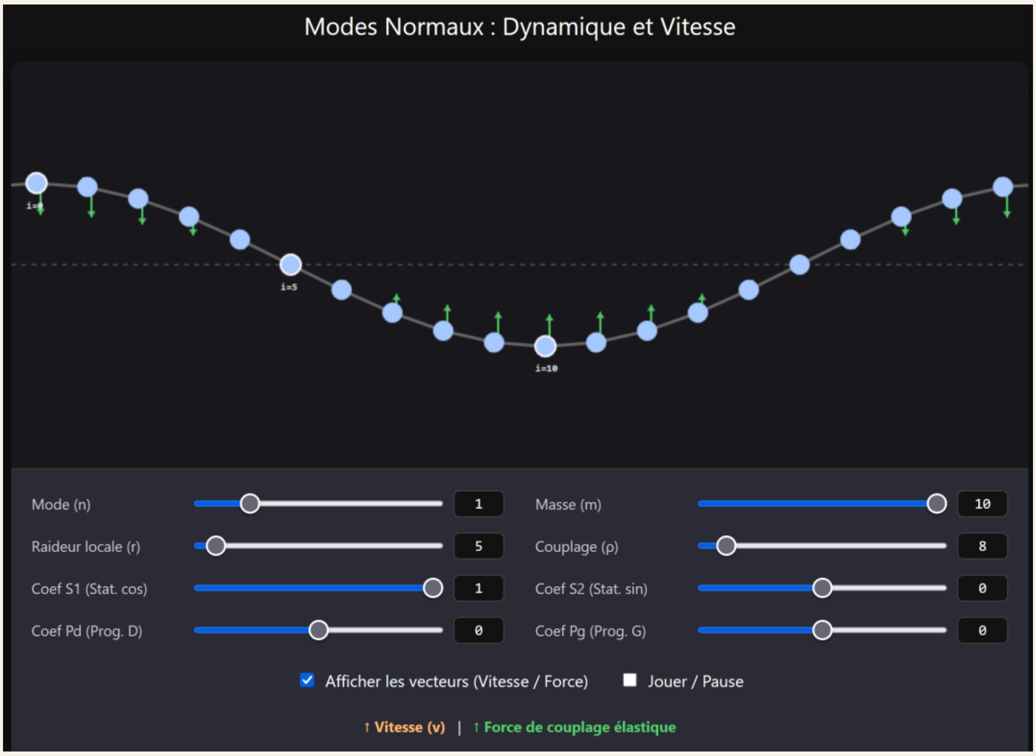
A. Cosine standing mode:

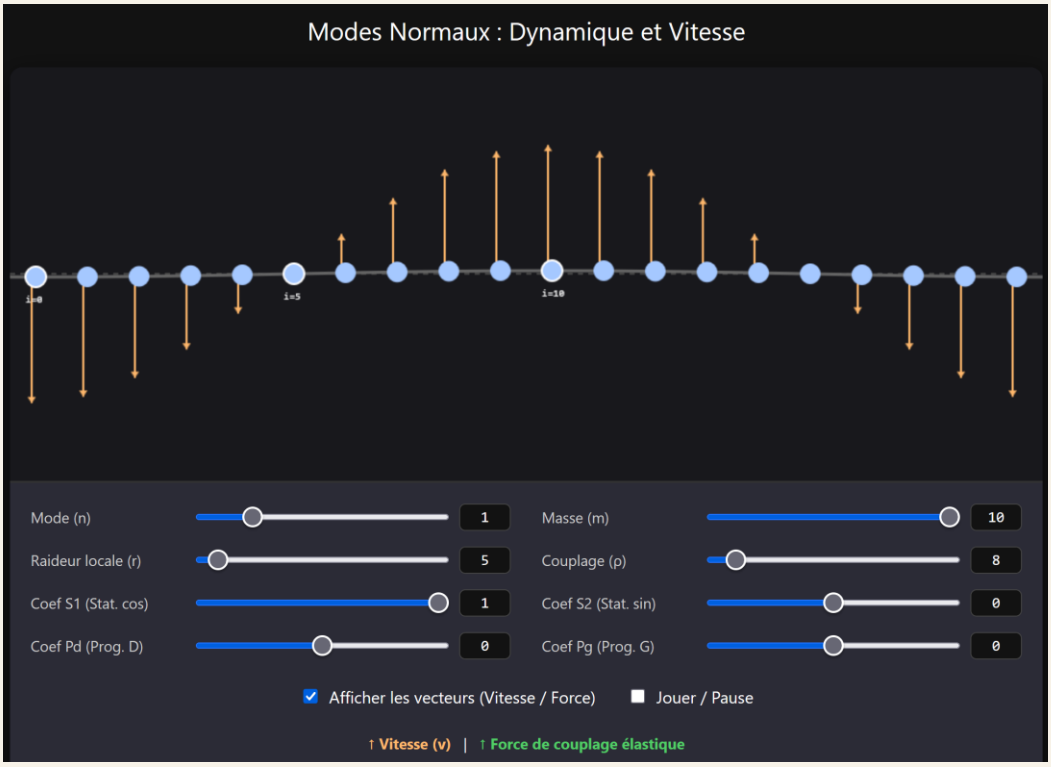
Let us begin with the mode known as the 'cosine standing mode'. In this mode, l(x,t) and v(x,t) are both cosines in their spatial dependence. The two functions therefore share the same spatial phase.

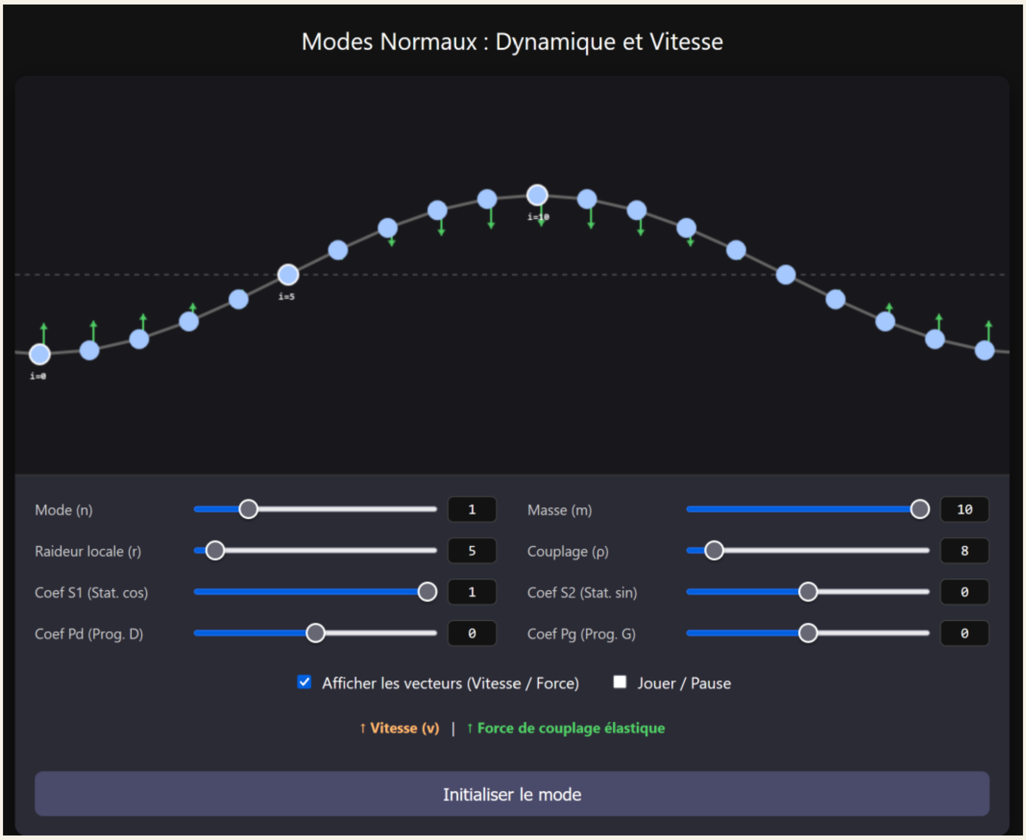
We can picture the evolution of l(x,t), shown below in blue, by adding a small-amplitude function with the same phase and wavelength, representing the velocity in orange. The resulting l(x,t+dt) is shown in green:

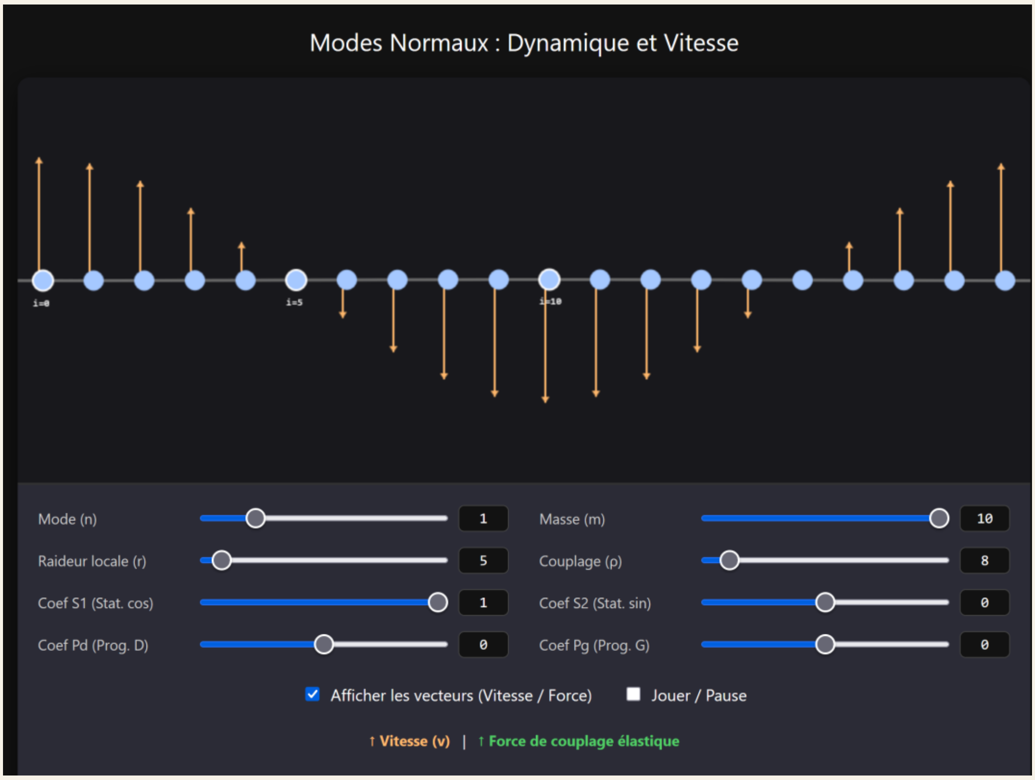


The key idea is that there are nodes: points in space at which the amplitude is always zero. Points for which l(x,t)=0 remain at zero because the velocity is also zero there. Recall that, for any given spring, velocity follows displacement with a phase lag. If both vanish at the same time, the spring is at rest at equilibrium. There are also antinodes, the parts of the wave at which the amplitude is always at its spatial maximum. The antinodes oscillate, but they always remain the points of greatest amplitude.


Modes Normaux : Dynamique et Vitesse
i=0
i=5
i=10
Mode (n)
1
Masse (m)
10
Raideur locale (r)
5
Couplage (p)
8
Coef S1 (Stat. cos)
1
Coef S2 (Stat. sin)
0
Coef Pd (Prog. D)
0
Coef Pg (Prog. G)
0
Afficher les vecteurs (Vitesse / Force)
Jouer / Pause
↑ Vitesse (v) | ↑ Force de couplage élastique

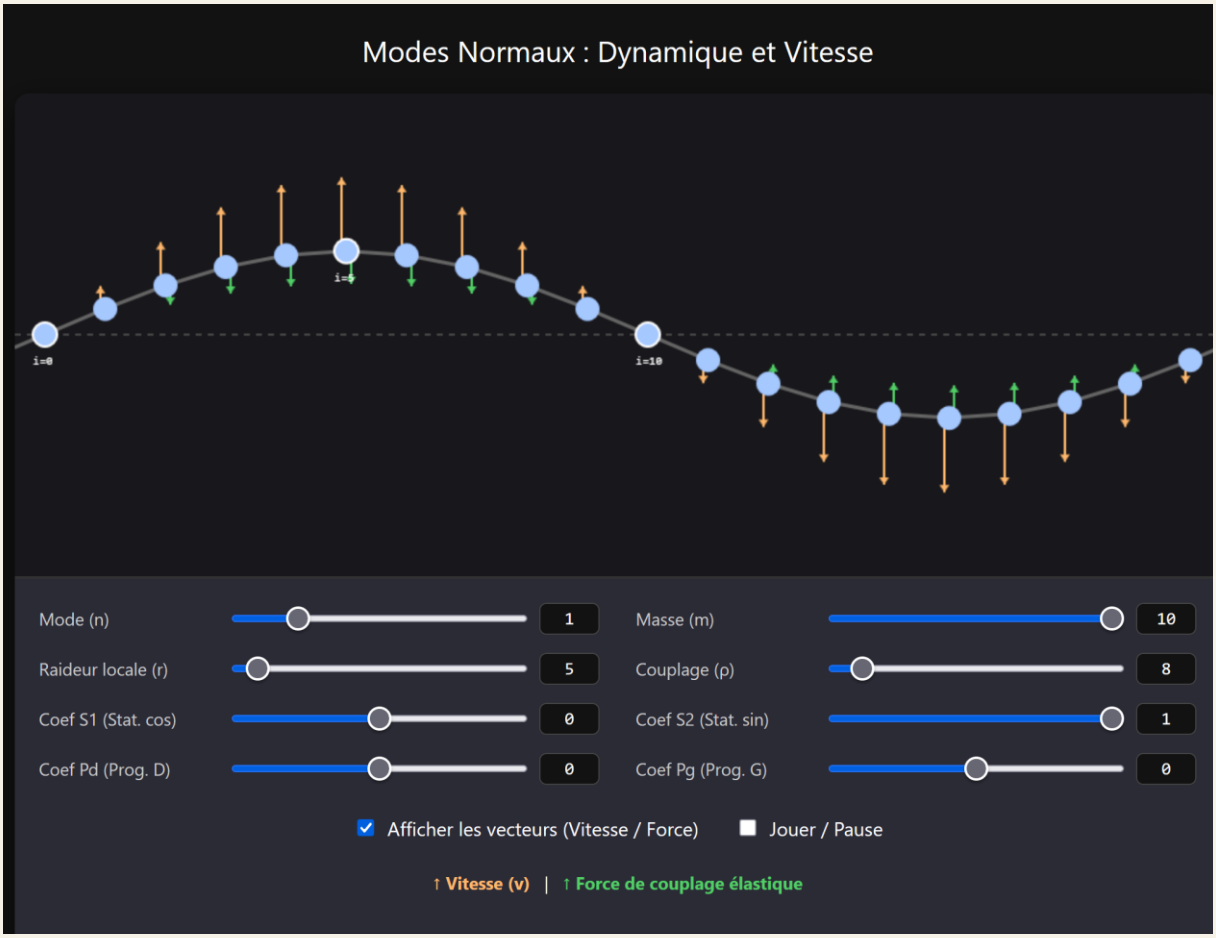

Modes Normaux : Dynamique et Vitesse
i=0
i=5
i=10
Mode (n)
1
Masse (m)
10
Raideur locale (r)
5
Couplage (p)
8
Coef S1 (Stat. cos)
1
Coef S2 (Stat. sin)
0
Coef Pd (Prog. D)
0
Coef Pg (Prog. G)
0
Afficher les vecteurs (Vitesse / Force)
Jouer / Pause
↑ Vitesse (v) | ↑ Force de couplage élastique





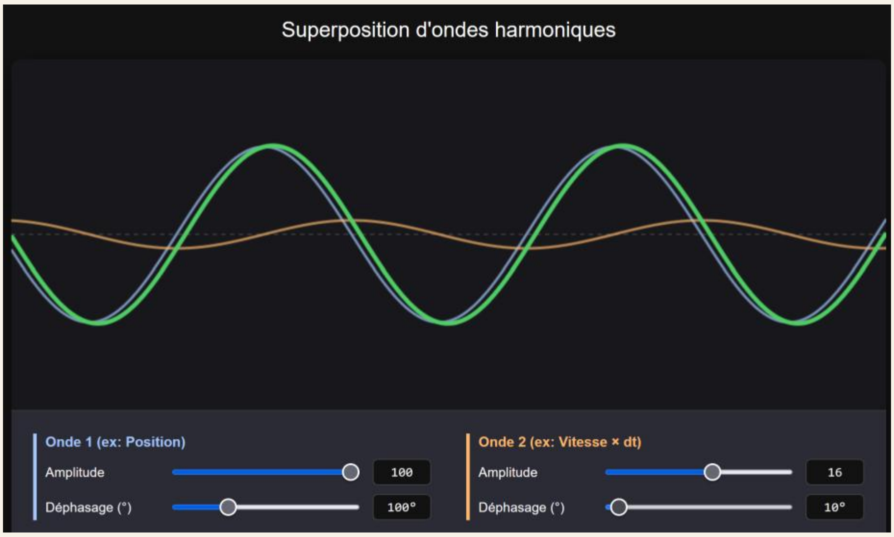
B. Sine standing mode:

There is little more to add. The principle is the same as for its cosine counterpart, except that the two spatial functions are now sines. The only difference is the location of the nodes and antinodes.

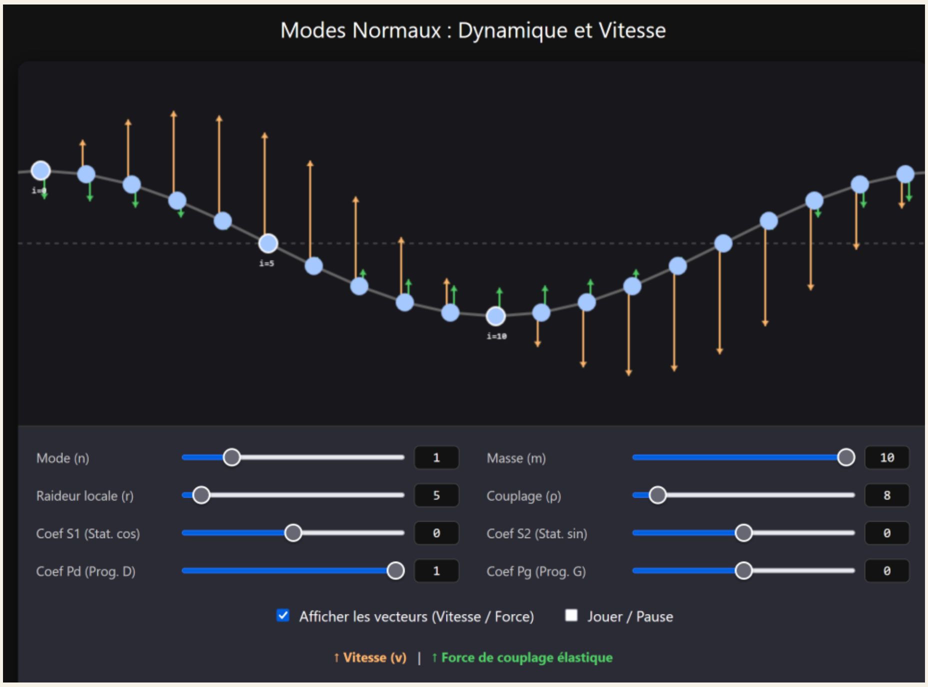


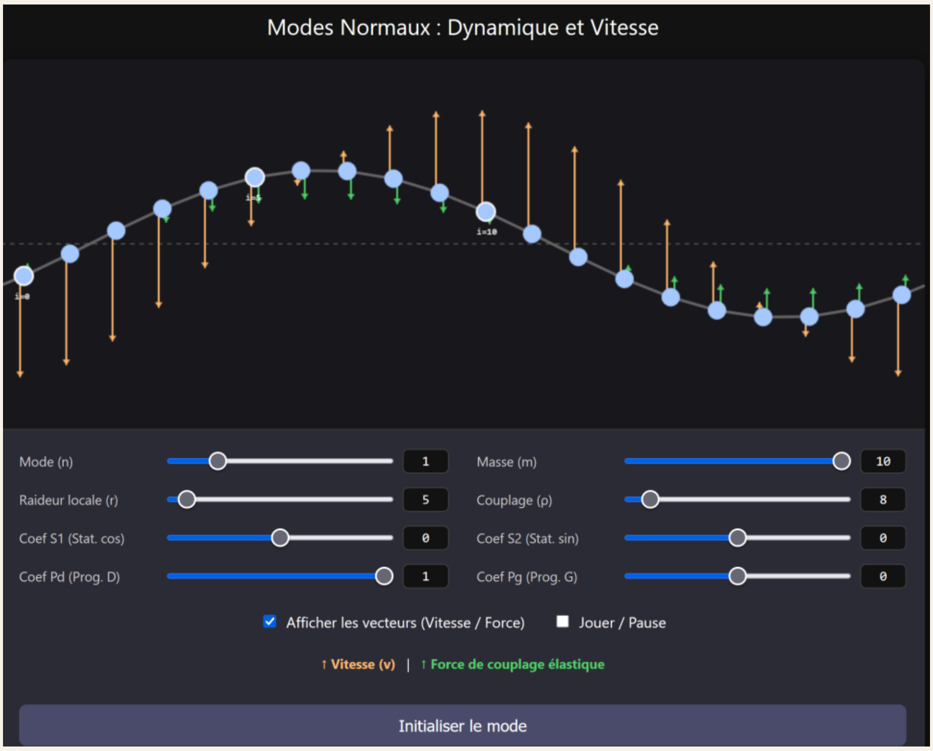
C. Right-moving travelling mode:

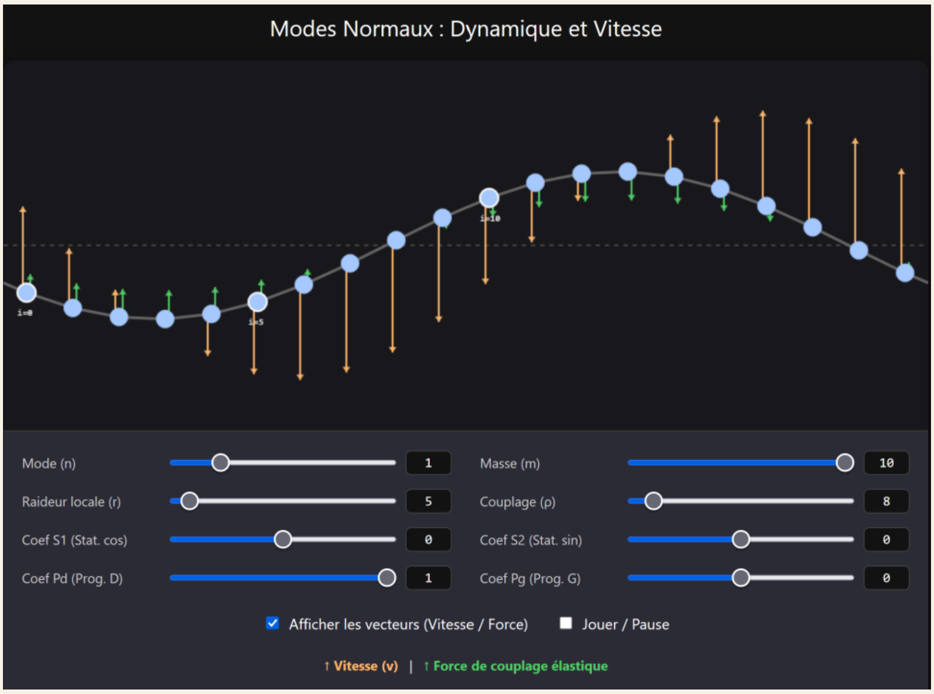
The other two modes are what we call travelling waves. They will be of particular interest because, in quantum field theory, they represent particle states with a specified momentum. They correspond to a spatial phase shift of π/2 between displacement and velocity. To see what this means physically, we again add a wave representing displacement as a function of x to a wave representing velocity shifted by −π/2. At a spatial crest the velocity is zero. Moving slightly to the right of the crest, the velocity becomes positive; moving slightly to the left, it becomes negative. At the next instant, the crest therefore shifts to the right: the springs to its right move upwards while the spring at the crest begins to move downwards under the forces acting on it. The entire wave structure translates to the right. One detail about the diagram should be noted. The green curve has a slightly greater maximum amplitude than the blue curve. This is an artefact of using a finite rather than genuinely infinitesimal increment so that the effect remains visible. The important feature is that the rightward displacement is far more pronounced than the increase in amplitude. In other words, it is precisely the spatial phase shift of π/2 between displacement and velocity that gives the wave its travelling character. By contrast, standing modes, for which this spatial phase shift is zero, do not propagate: the wave merely 'breathes' in place.



The propagation of a right-moving travelling wave is shown below:

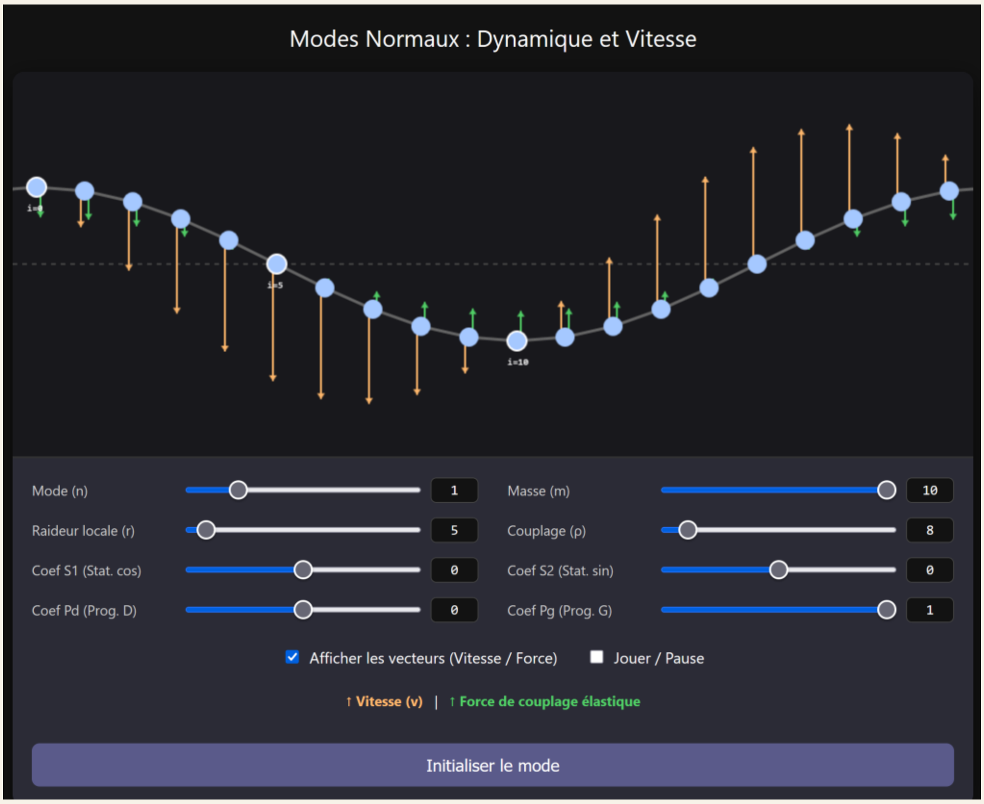






D. Left-moving travelling mode:

The principle is the same, but with a phase shift of +π/2, which produces propagation to the left.

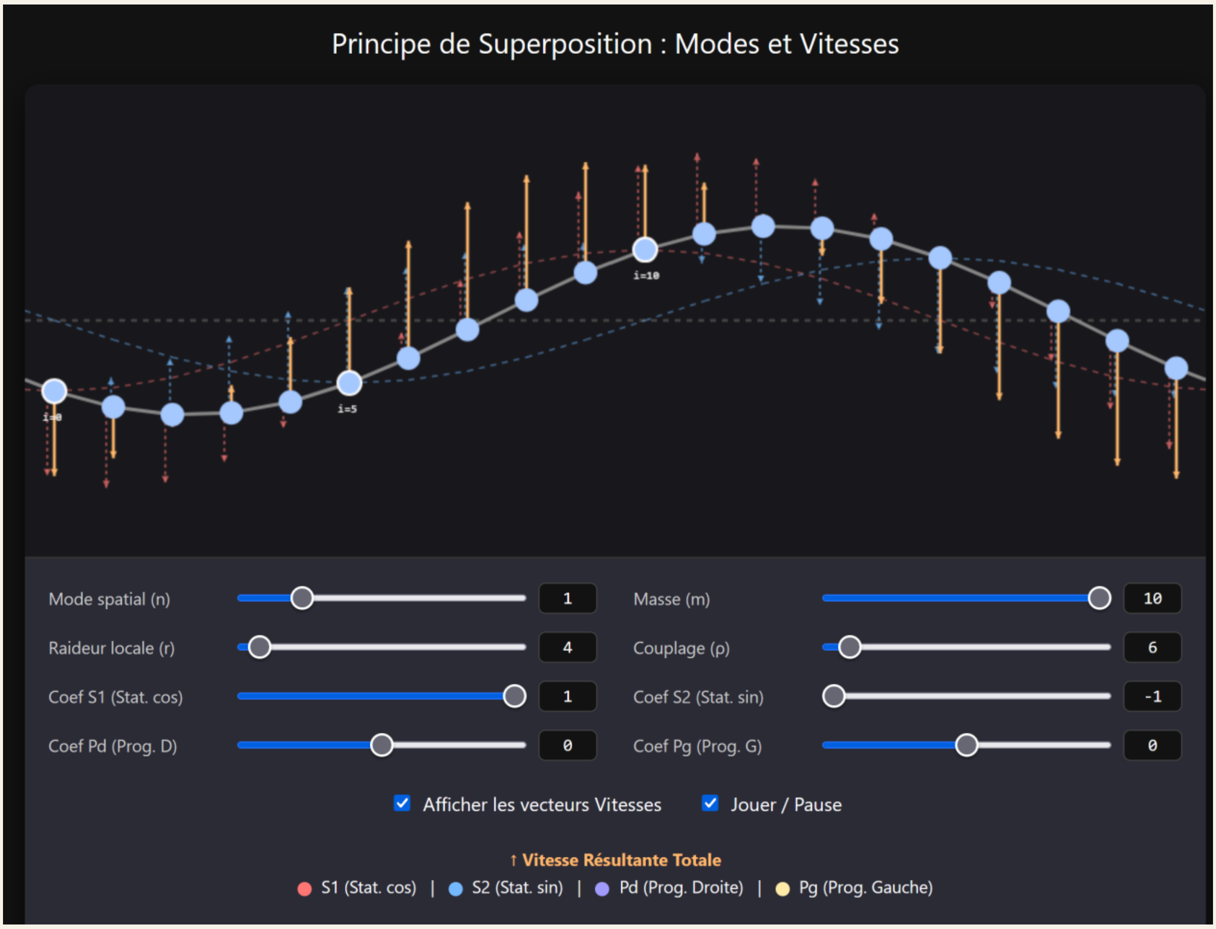


II.5 Basis Solutions

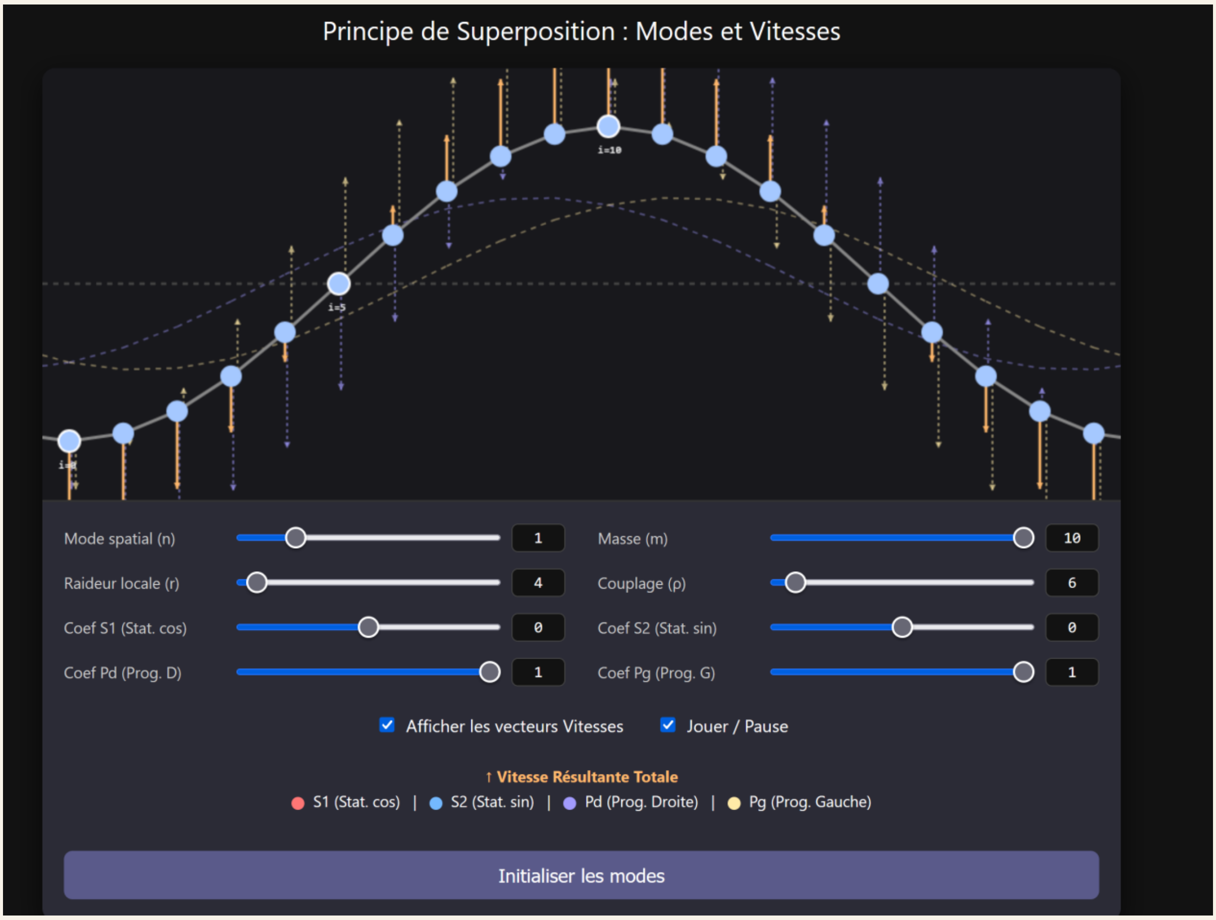
A. Adding and Subtracting Modes

We now come to something that will have very deep consequences later on. What happens if we take the sine standing mode and the cosine standing mode, and for each spring add together the positions and velocities of the two modes? In the diagram below, I have superposed the cosine standing mode at equal amplitude (red velocity vectors and a red dashed spatial displacement profile) with the sine standing mode (blue velocity vectors and a blue dashed spatial displacement profile). If, for each spring, we add the displacements of the two modes and also add their velocity vectors, we obtain respectively the orange velocity vectors and the positions of the masses. What we see is that adding these two modes produces the right-moving travelling mode.

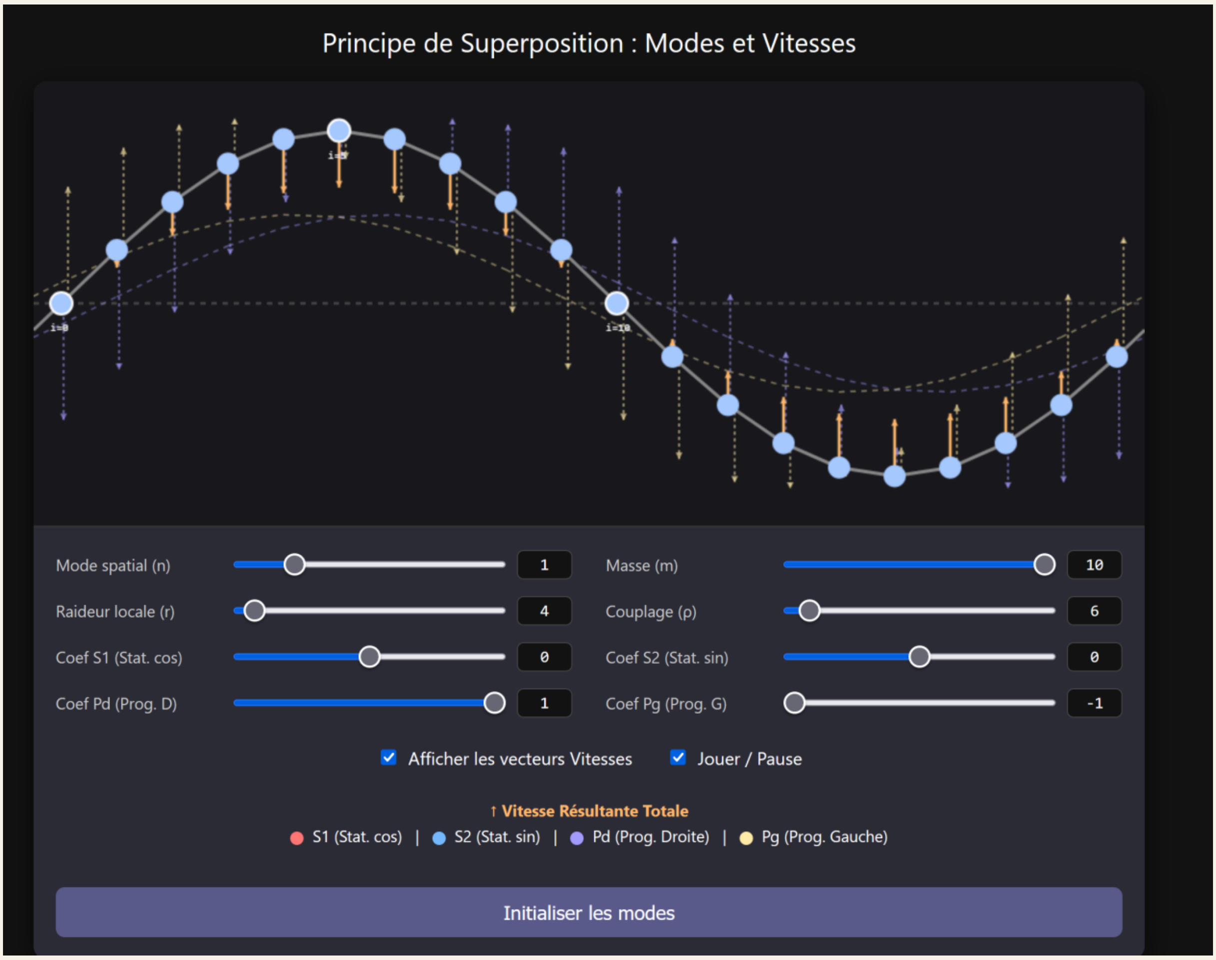
If we now take the difference between these two modes, we obtain… the left-moving travelling mode.



And what happens if we add the two travelling modes? Take a guess:



We obtain the cosine standing mode. And what if we take the difference between the travelling modes?



We obtain, of course, the sine standing mode.

(see Interactive Animation 8)

For convenience, let us give these modes short names: S1 for the sine standing mode, S2 for the cosine standing mode, Rd for the right-moving travelling mode, and Ld for the left-moving travelling mode. One important clarification is needed to make the following relations exact: S2 must have a temporal dependence cos(ωt), corresponding to maximum displacement at t = 0, whereas S1 must have a temporal dependence sin(ωt), corresponding to zero displacement at t = 0. These two standing modes are physically equivalent; they are merely shifted by a temporal phase of π/2. With these conventions, S1 + S2 = Rd and S2 - S1 = Ld are exact equalities, while Rd + Ld = 2·S2 and Rd - Ld = 2·S1.

We therefore have:

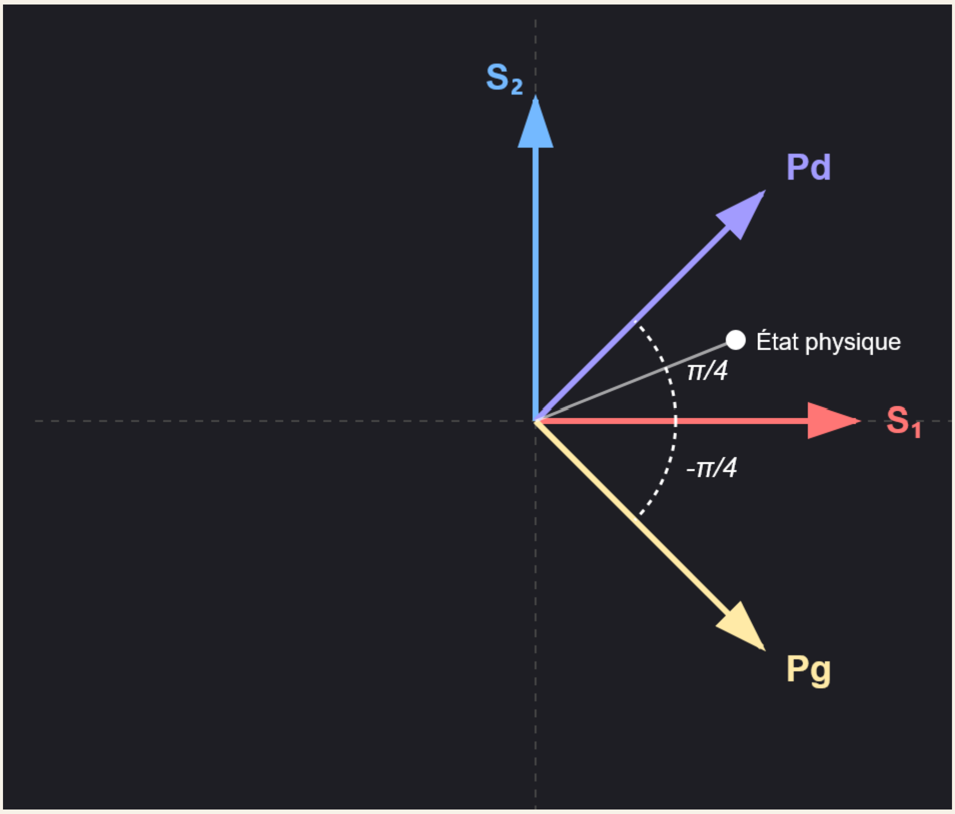
$$S1 + S2 \propto Rd \qquad\qquad S1 - S2 \propto Ld$$

$$Rd + Ld \propto S2 \qquad\qquad Rd - Ld \propto S1$$

B. The Vector Space of Solutions: A First Approach

We shall soon see that the relations among $S1$, $S2$, Rd and Ld are the same as the relations among four vectors in a two-dimensional space: $\overrightarrow{S1}$ and $\overrightarrow{S2}$ form one orthonormal basis, while $\overrightarrow{Rd}$ and $\overrightarrow{Ld}$ form another orthonormal basis rotated relative to the first by $\pi/_4$. We shall even see that a norm and an inner product can be defined for them, just as for ordinary vectors. Although this may seem very abstract at first, all the solutions of mode *n=1* mathematically form a two-dimensional vector space. This means that every solution belonging to mode n = 1 can be represented by a vector in that space. It can therefore be expressed either as a linear combination of $S1$ and $S2$, or as a linear combination of Rd and Ld.

Choosing to express the solution as a combination of standing waves or as a combination of travelling waves is entirely equivalent to choosing a basis in the corresponding vector space.



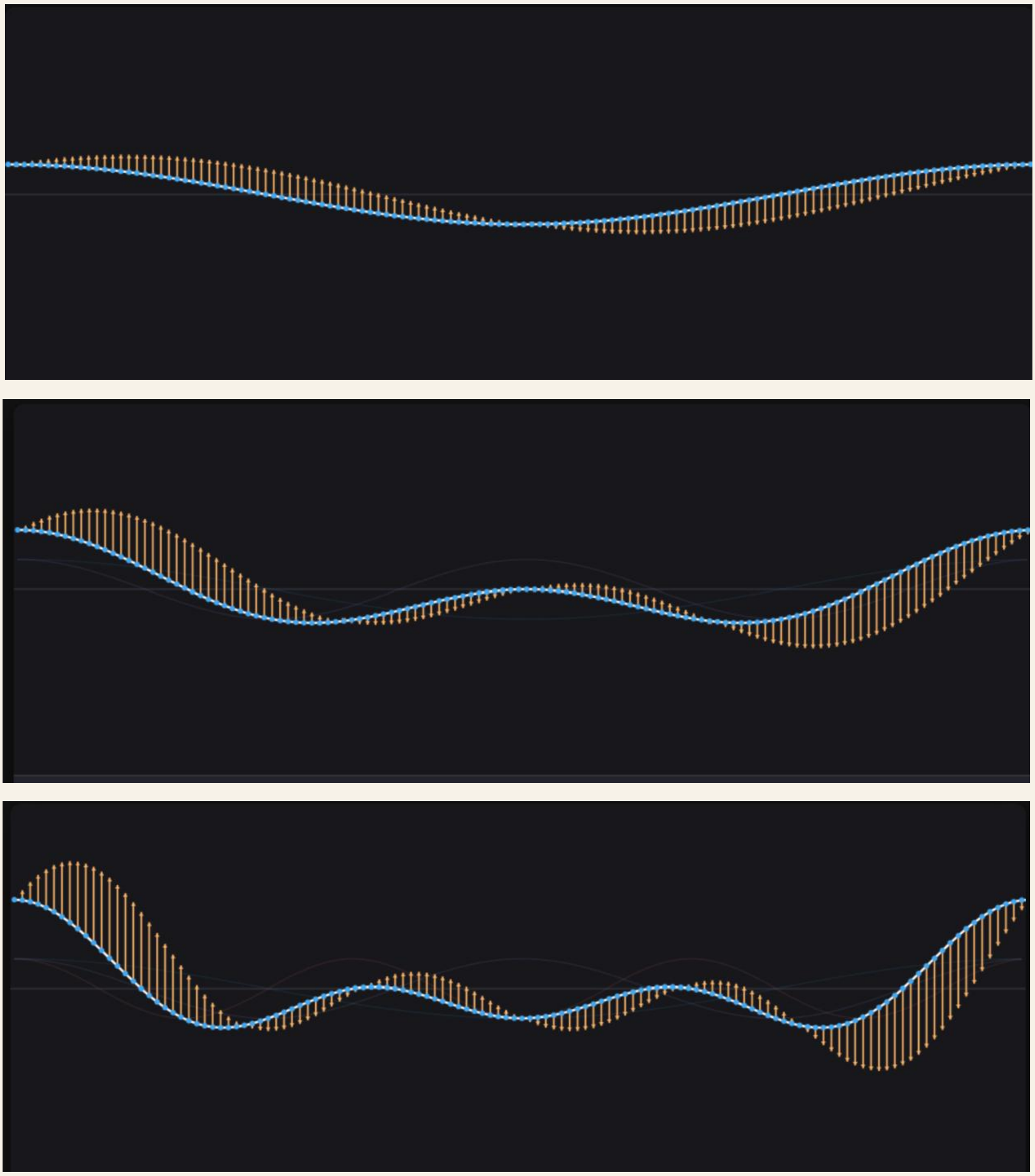
The diagram shows that adding the vectors Ld and Rd gives a vector proportional to S1; subtracting Ld from Rd gives a vector proportional to S2; adding S1 and S2 gives a vector proportional to Rd; and subtracting S2 from S1 gives a vector proportional to Ld. We therefore recover exactly the relations among the four solutions established above. Later, we shall see that an operation can be defined between these four wave solutions that produces exactly the same results as the inner products between the corresponding vectors.

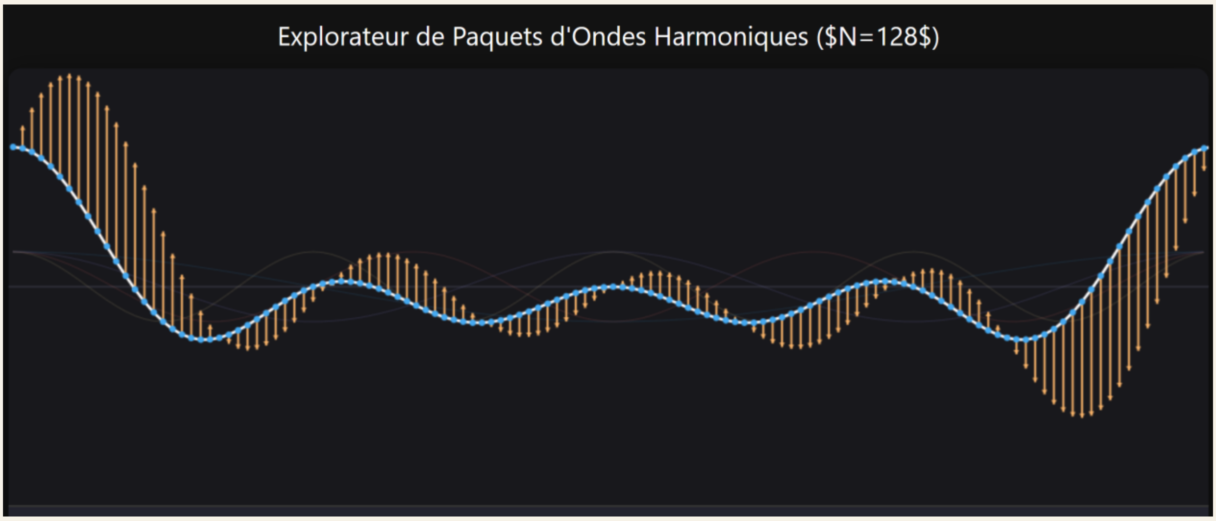
The same is true for modes n = 2, 3, and so on without limit. For the same reasons, each of these modes forms its own two-dimensional vector space. Taken together, the full solution space is infinite-dimensional. None of this is strictly necessary for understanding the main ideas developed in the first parts of this series. I introduce it now because revisiting the same structure from several different angles whenever the opportunity arises gradually builds an understanding of an aspect of modern physics that even some advanced students struggle to grasp deeply. These considerations become indispensable in quantum theory. In quantum mechanics, states correspond to vectors that can be represented by wavefunctions analogous to the waves we have just studied. The essential difference is that the relevant vector space is complex: the components of the vectors in the various possible bases are complex numbers rather than real numbers. Paradoxically, we shall see that this makes matters simpler rather than more complicated.

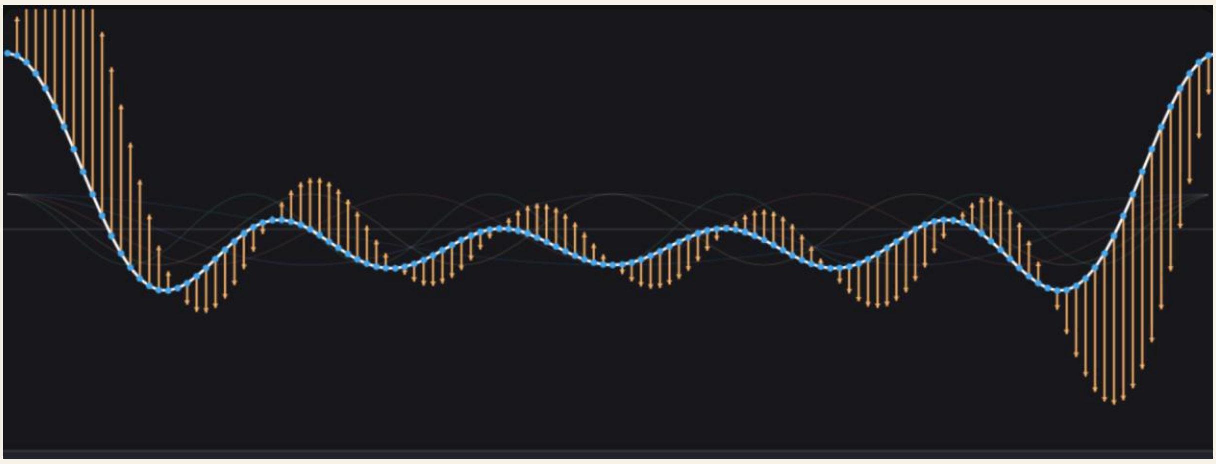
II.6 Coherent Superpositions of Waves with Different *n* and Wave Packets

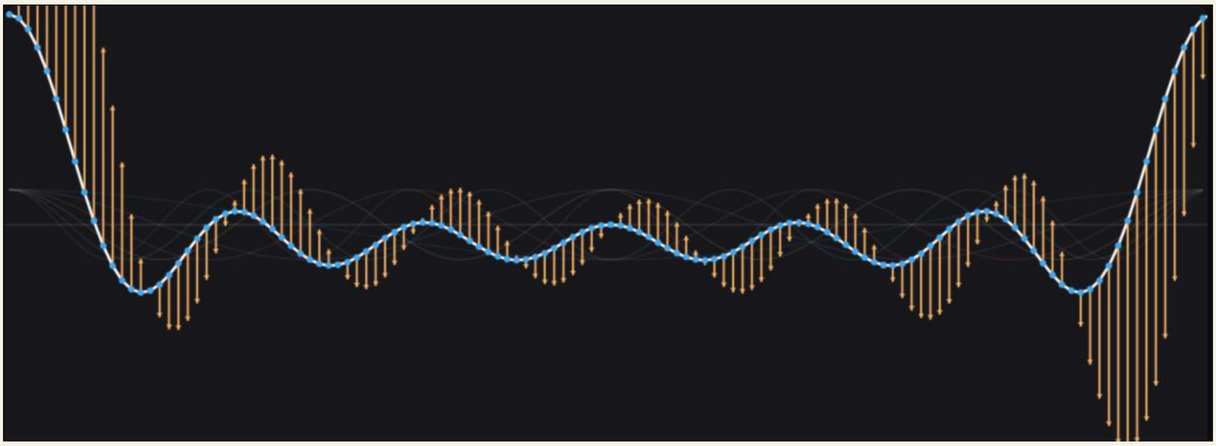
Instead of superposing waves with the same n, we shall now superpose waves with different values of n. We shall use only right-moving travelling waves. We begin with a pure n = 1 wave, and in each successive image add the next mode: n = 2, then 3, and so on up to 15. Every mode has the same amplitude. We define the zero of phase for all modes at x = 0, on the far left, so that all the modes are in phase at that point in space. This is called a coherence condition. Before looking at the images, we can predict what will happen. At x = 0, all the modes we add are at their maximum and in phase, so their amplitudes simply add. The oscillator at x = 0 will therefore very quickly acquire a very large amplitude. The lowest modes vary slowly in space and will still have large values at x = 1, 2, 3, and so forth. Because the boundary conditions are periodic, the same is true near the oscillators at the far right. Throughout this region, where the phases of the waves remain aligned, the amplitudes add directly. This is constructive interference. Farther away from the oscillators surrounding x = 0, however, the phase differences between the waves become larger. In some regions, certain modes are even in opposite phase and cancel one another: this is destructive interference. More generally, the farther one moves from the region of constructive interference, the more desynchronised the phases become and the more the interference terms tend, on average, to cancel. As more high-wavenumber modes are added, the region in which the modes remain in phase becomes increasingly narrow. We shall therefore see a progressively smaller region acquire an increasingly large amplitude through constructive interference, while elsewhere the modes are out of phase and the total amplitude tends to become smaller, apart from some spatial fluctuations.

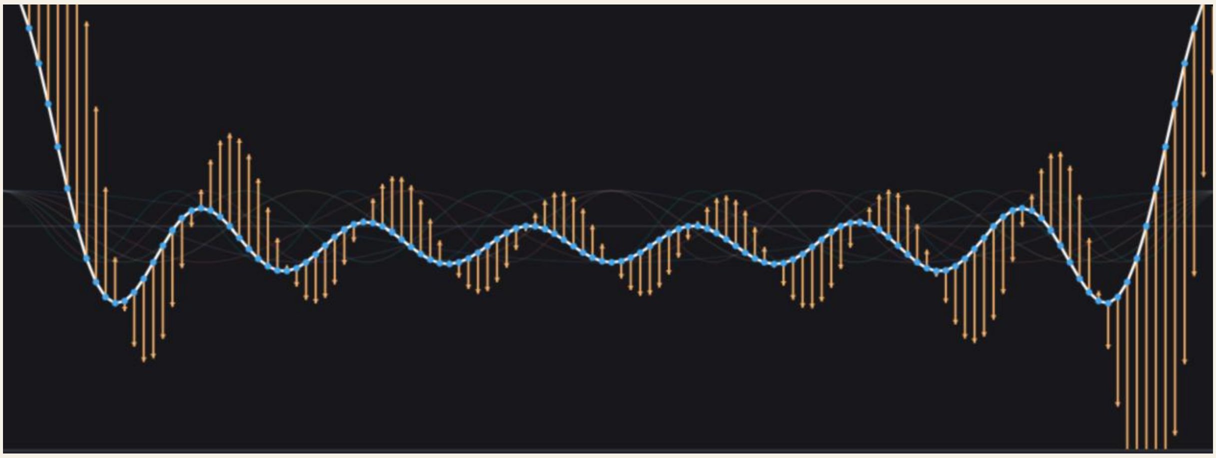
(see Interactive Animation 9)

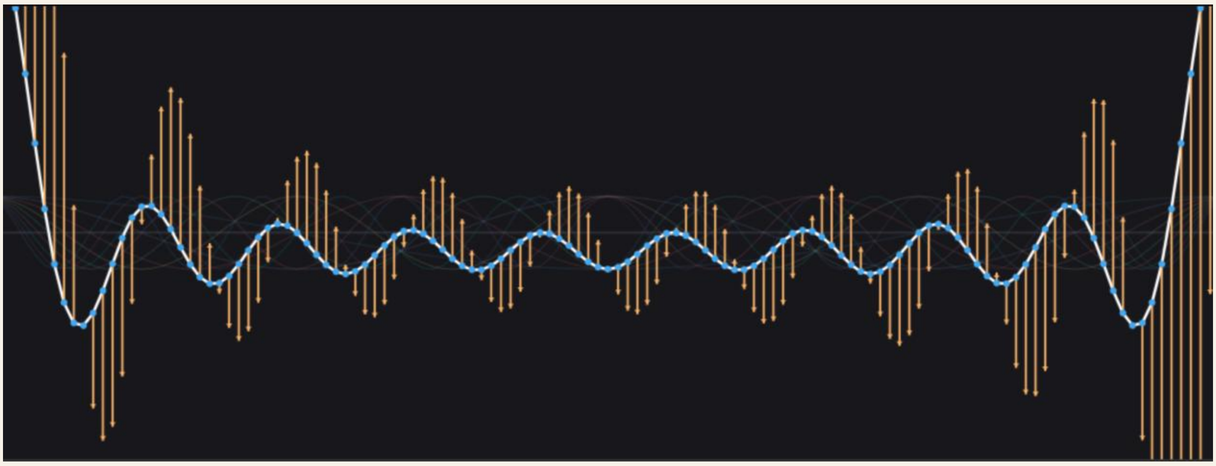


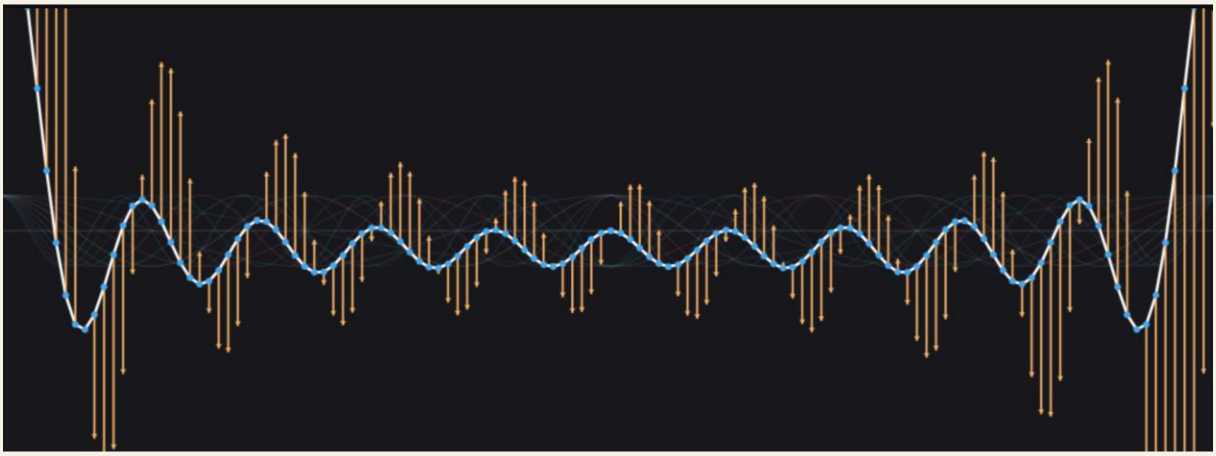

Explorateur de Paquets d'Ondes Harmoniques ($N=128$)

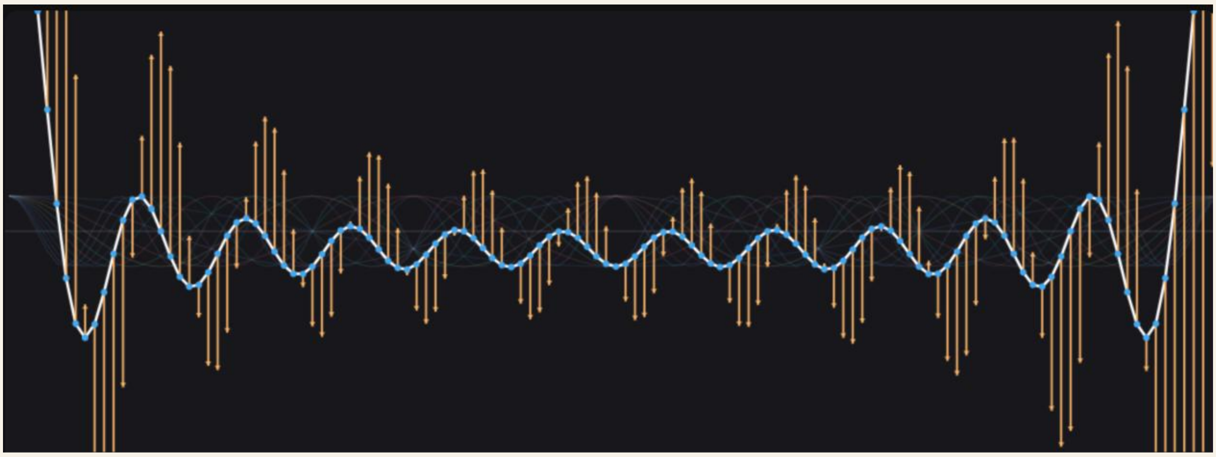


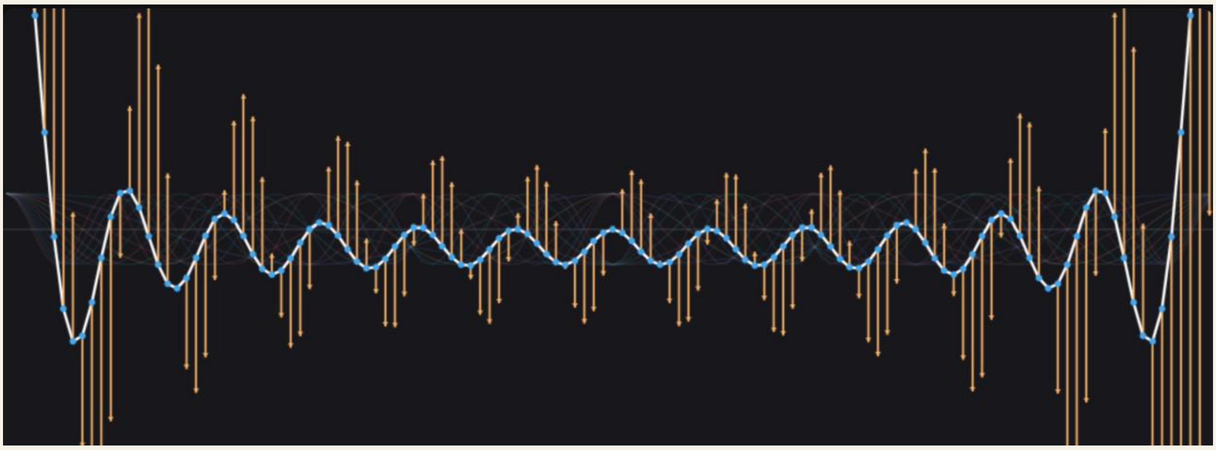


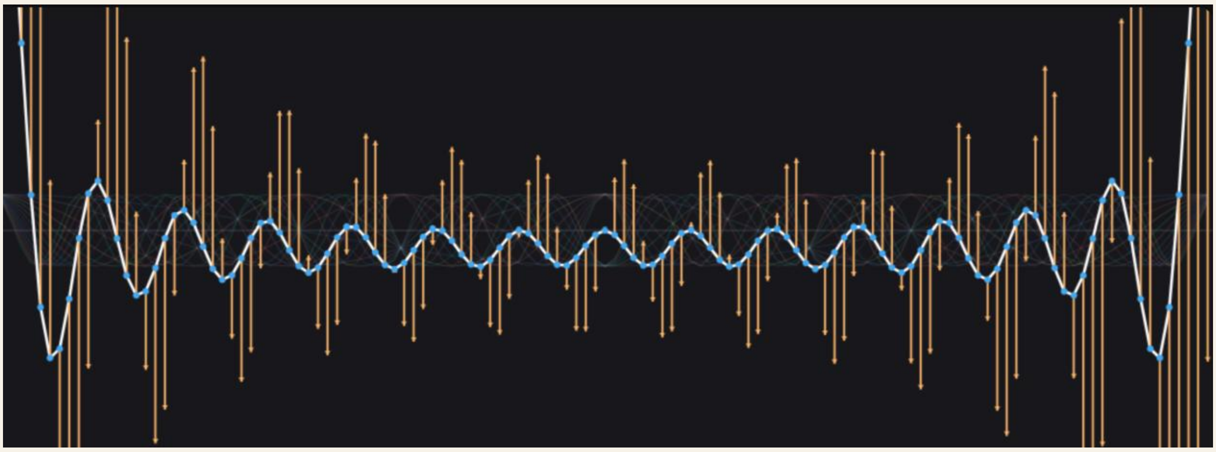


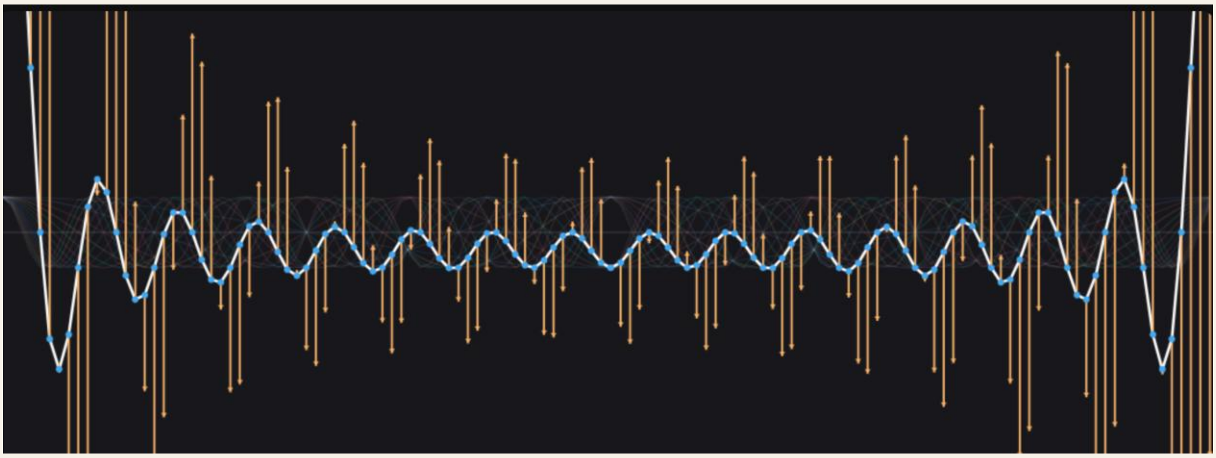


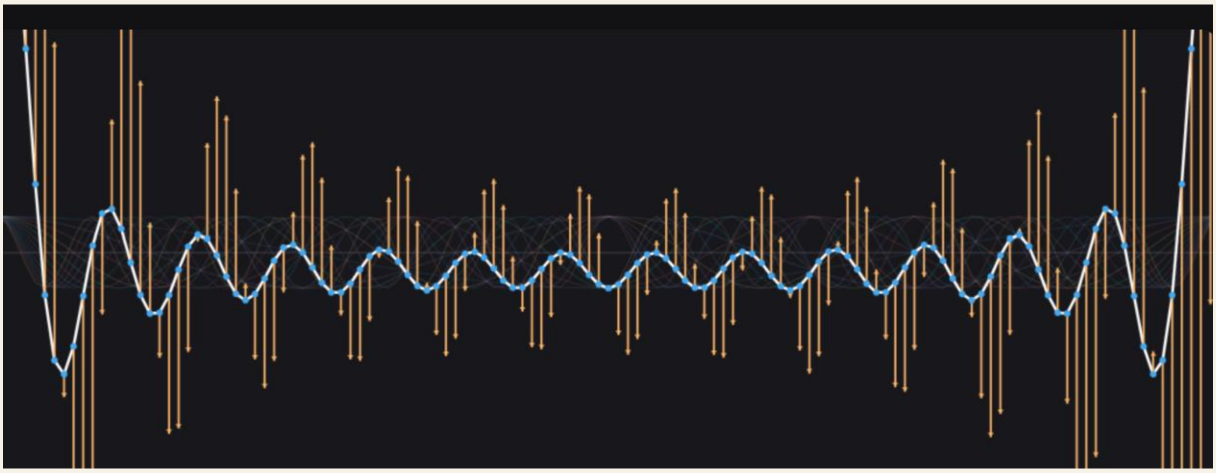












The high-amplitude region in which all the modes are in phase is called a wave packet. In the classical field-based approach developed here, this kind of structure will play the role of a particle. We shall see that, under certain conditions, wave packets can follow trajectories very close to those of Newtonian or even relativistic particles. The latter case will emerge in the next section when we introduce the Klein-Gordon equation.

II.7 Loss of Coherence, Dispersion, and White Noise

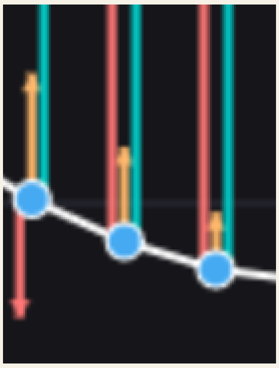
Unless the restoring constant of the individual springs is zero, we shall soon see that the different modes do not all travel at the same speed. The maxima of all the waves that initially coincide at the centre will therefore gradually drift apart. Their phases will cease to remain aligned, and the coherent wave packet will spread out over space. This phenomenon is called dispersion. The packet does not disappear because its energy is lost; rather, the different modes that compose it progressively dephase relative to one another. The initially localised structure therefore becomes broader and less sharply defined.

Once coherence has been lost, the spatial pattern becomes increasingly irregular. We may call the resulting state a white wave, by analogy with white light, which is composed of many different frequencies whose phases are not mutually coherent. The comparison is not exact, but it captures the essential point: when many modes with different frequencies are superposed without stable phase relations, no persistent large-scale structure emerges from their combination.

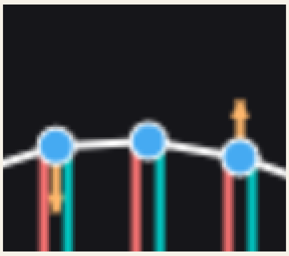
II.8 Wave Packets in an Inhomogeneous Medium: Nested Harmonic Oscillators

A. Reflection: When a Wave Bounces Back

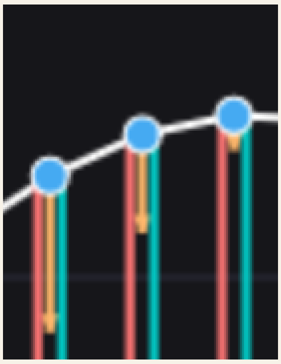
We shall now modify the boundary conditions of our spring system. From now on, the spring at each end will be connected by an elastic coupling to a fixed wall. Let us consider how a wave behaves as it approaches such a boundary. During the ordinary propagation of a travelling wave, a spring moving upwards pulls the next spring upwards through the elastic coupling shown in green. At the same time, it is itself pulled by the preceding spring, as illustrated below.



The central spring in the figure then reaches the crest of the wave and begins to experience a downward force from its two neighbours.

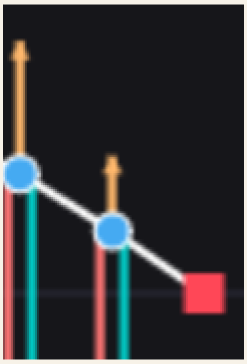


Next, the spring on the right rises above the central spring.

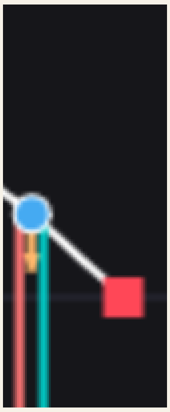


Throughout this process, the spring-velocity vectors shown in orange point in the direction opposite to the height gradient. This relation indicates the direction of propagation of the wave: towards the right.

When the wave reaches a boundary, the wall is not carried along by the final spring. It therefore continues to exert a substantial force on that spring.



Because this downward force continues to act, the velocity vectors reverse before the height gradient itself has changed direction.

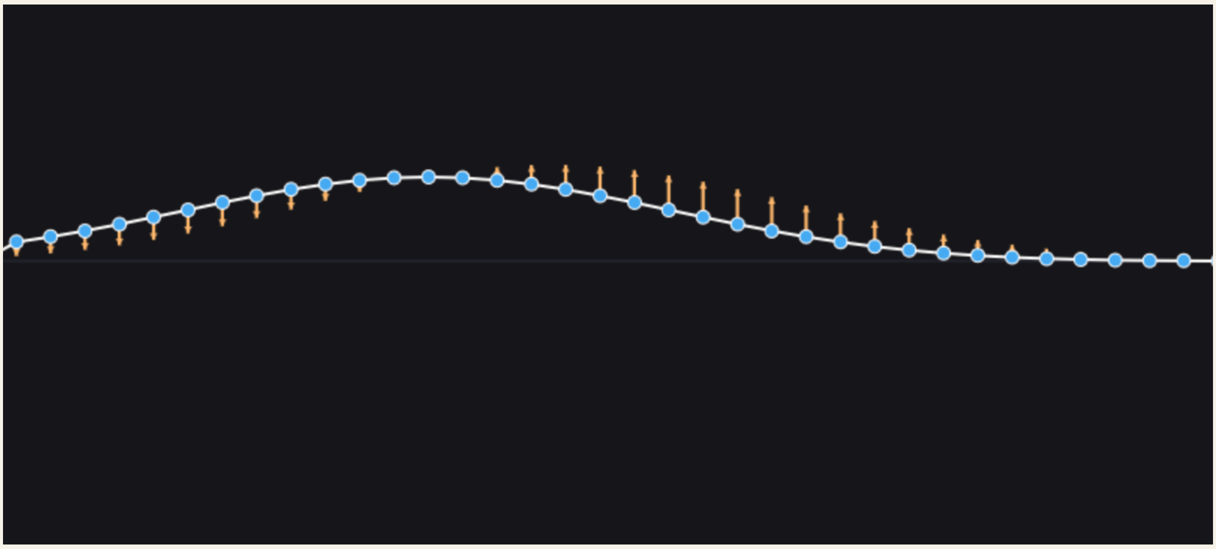


The velocity vector therefore becomes aligned with the gradient, marking a reversal in the direction of wave propagation. The wave appears to bounce off the wall and travel back towards the left. Individually, every spring continues to oscillate more or less regularly up and down. Collectively, however, the synchronisation between velocities and heights is reorganised, with the result that a structure that previously appeared to translate to the right now translates to the left.

(see Interactive Animation 10)

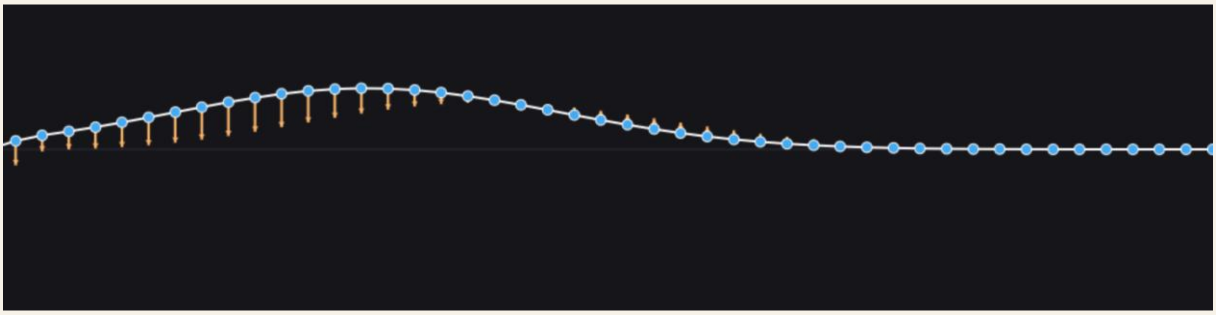
B. Turning Back: When a Wave Encounters a Potential Stronger Than It Is

We shall now observe something particularly interesting. We shall modify the springs in a subtle way by adding an additional local stiffness that increases linearly towards the right. In other words, the farther to the right a spring lies, the stiffer it becomes. Let us prepare an initial state in which a wave packet on the left begins to move towards the right. What happens?

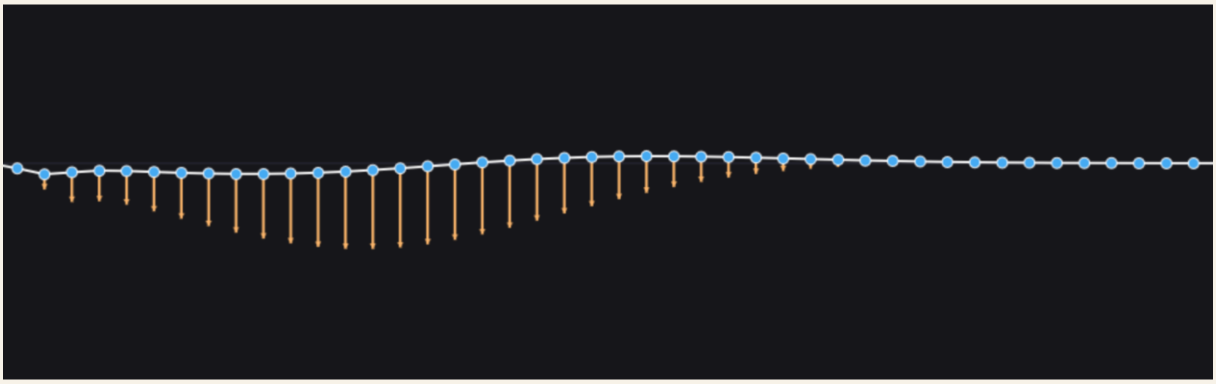


Initially, the orange velocities appear equally large—though oppositely directed—on the leading and trailing sides of the wave. The packet therefore seems to move uniformly to the right while preserving its structure.

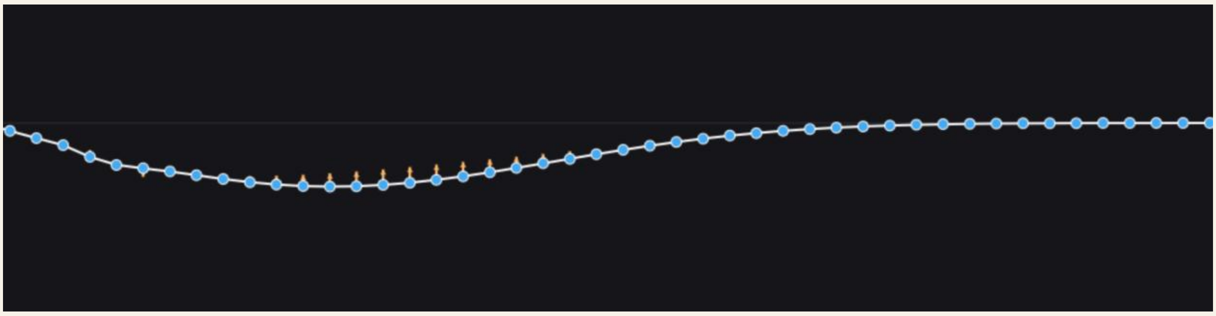
As it travels farther to the right, however, an increasing asymmetry develops between the velocities on the right and those on the left. The reason is that the springs become progressively stiffer towards the right and increasingly resist being displaced.

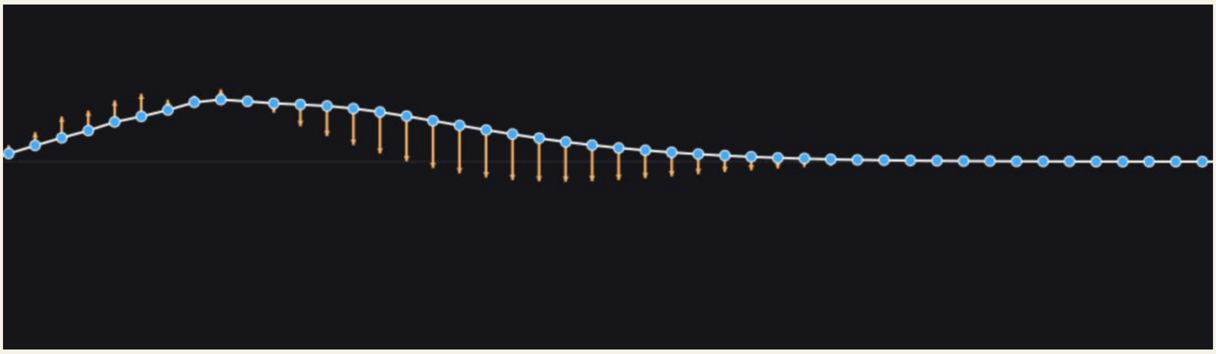


A point is eventually reached at which the springs are too stiff to be lifted any farther. Through inertia, the springs on the left then begin to dip downwards.



A new reversal between velocity and gradient begins to appear on the left-hand side of the wave. This indicates that part of the wave is being reflected and is beginning to travel back in the opposite direction.

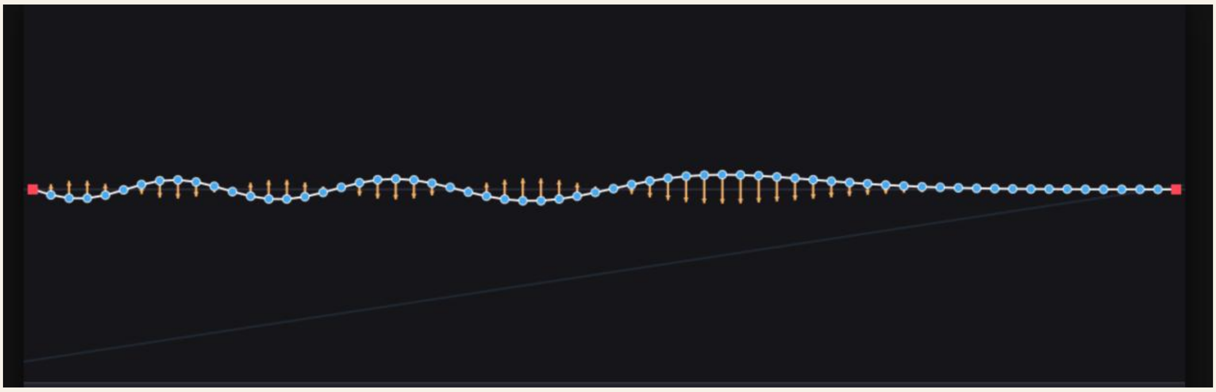




C. Steady Regimes and Potentials: Reconsidering Changes in the Direction of Waves

We shall now look for steady regimes of a wave moving through a potential that increases linearly towards the right. That is, we seek a wave whose frequency remains fixed everywhere and at all times.

If the frequency is the same for every spring, then the effective restoring-force coefficient must also be the same everywhere. We have seen that this effective restoring force is the sum of the local restoring term of each spring and the curvature term supplied by the elastic couplings. If the local restoring coefficient increases linearly towards the right, there is only one way for the effective coefficient to remain constant: the curvature contribution must decrease accordingly. In other words, since wavelength is inversely proportional to the square root of curvature ($\lambda \propto 1/\sqrt{\text{curvature}}$), the wavelength must increase towards the right.



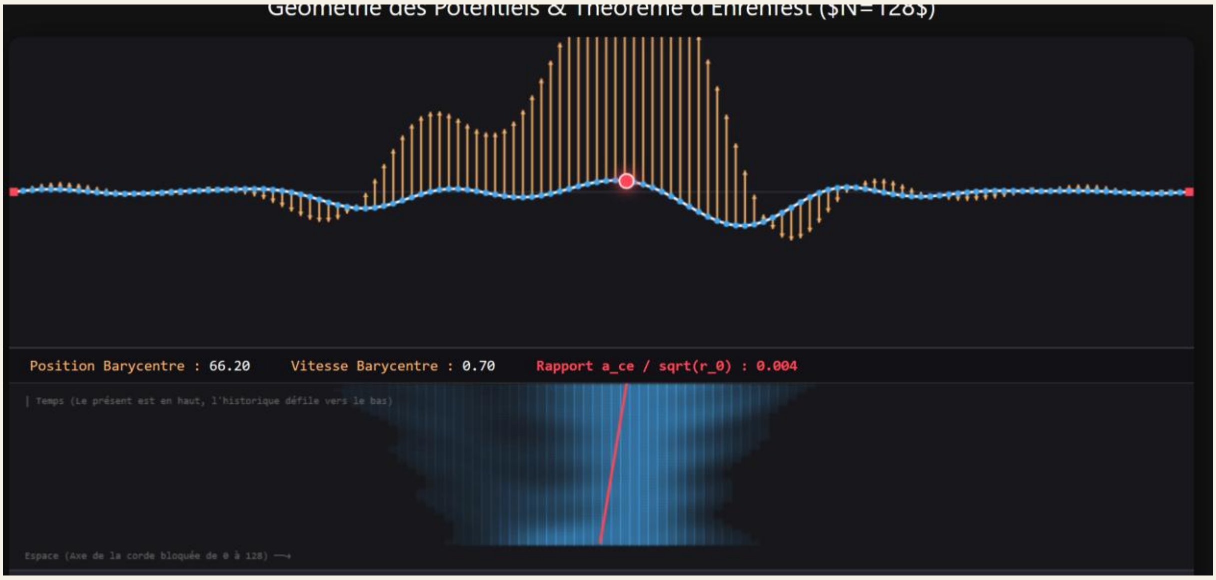
This result allows us to reinterpret the change in direction of travelling wave packets in a linearly increasing potential. As the waves move into regions of higher potential, their wavenumbers progressively decrease until they reach zero and then change sign. The point $k = 0$ at which the wave turns around is precisely the classical turning point of the WKB approximation—named after Wentzel, Kramers, and Brillouin—which we shall examine in Part 3.

D. Dynamics of Wave Packets

We can now assemble the pieces and witness our first example of emergence in action. Under certain conditions, wave packets display a dynamics resembling that of particles. To make the resemblance more explicit, we shall mark with a red point the barycentre of the energy carried by the springs. At every time step, we calculate the sum of the potential and kinetic energies of each spring and then determine the barycentre of this energy distribution. We shall call the motion of this barycentre the motion of the wave packet's centre of energy.

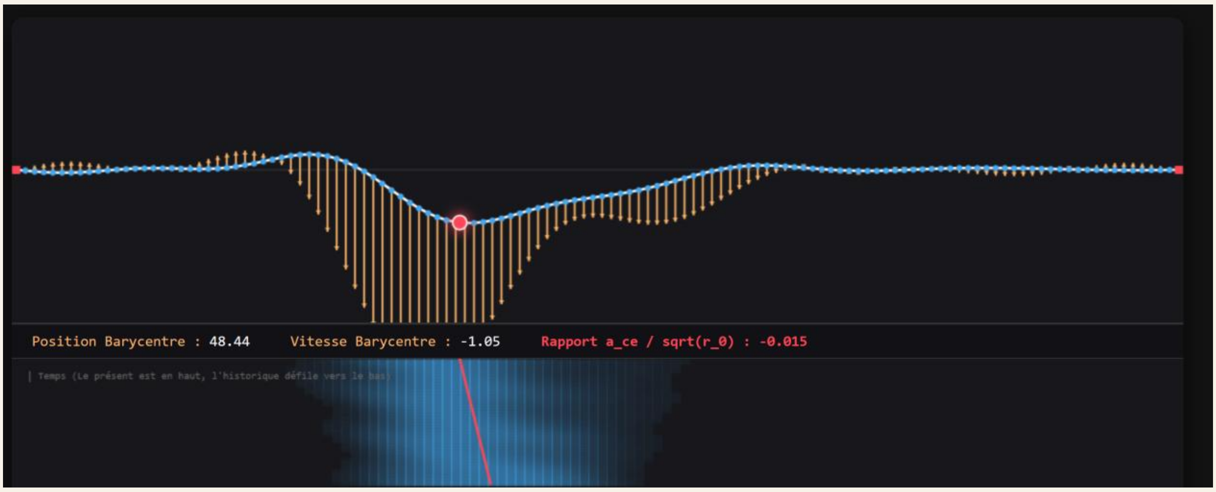
a. The Principle of Inertia and Elastic Collision with a Wall

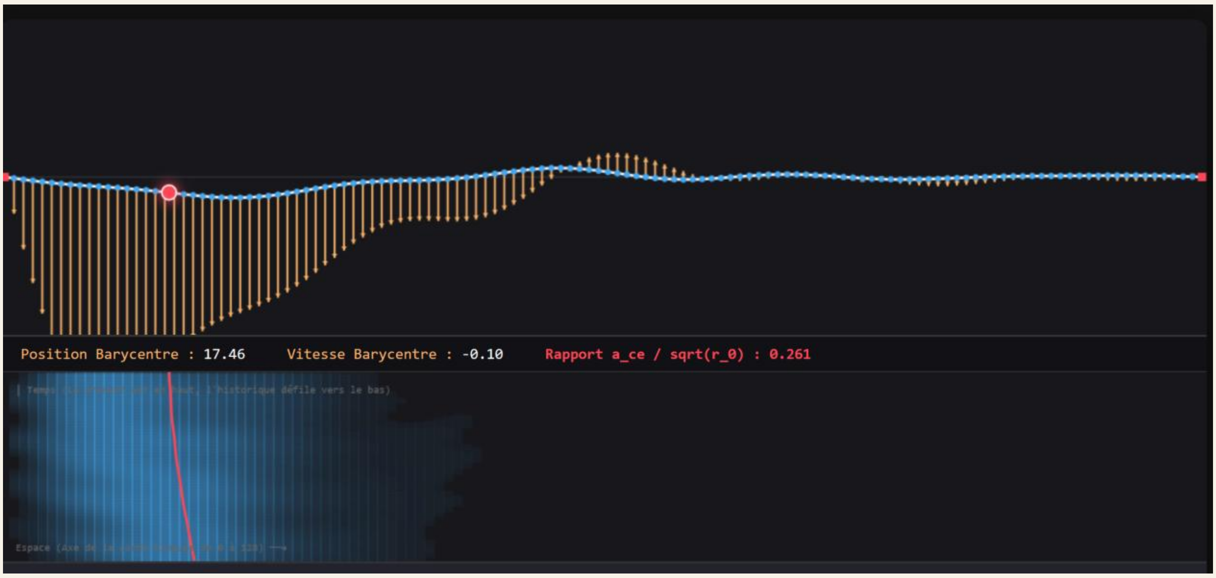
Let us first examine the centre of energy of a wave packet moving to the right through a homogeneous medium, far from the boundaries. Beneath the image of the packet, the trajectory of its centre is displayed. It is approximately a straight line, in accordance with the principle of inertia:

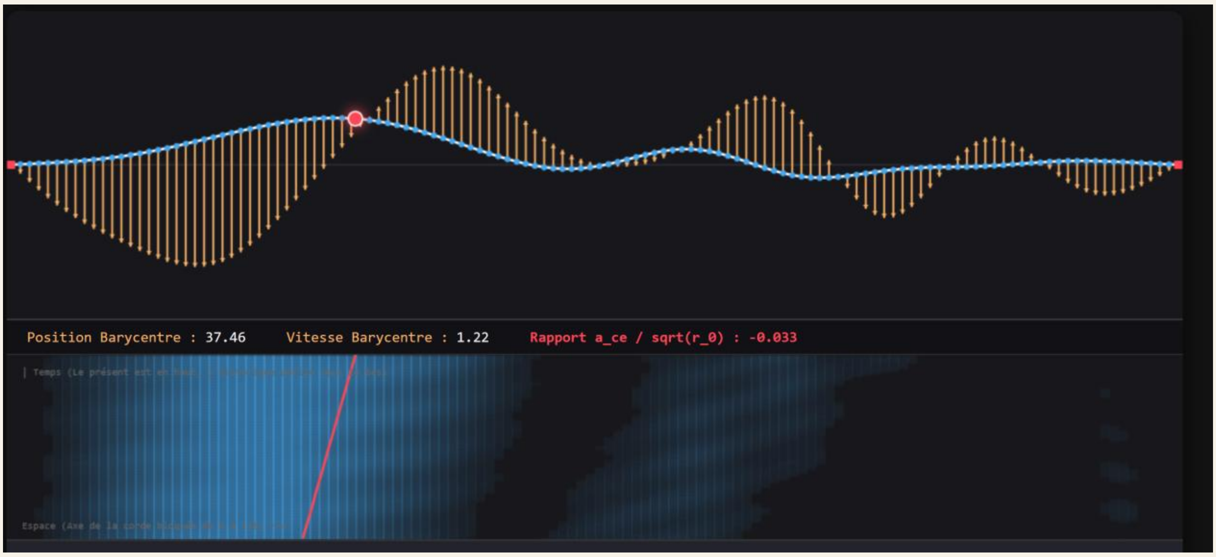


Tracking the velocity confirms that it remains within about 0.01 of the value 0.70, as long as waves reflected by the boundaries do not return and interfere with the packet's dynamics.

When the wave is sent towards a wall—shown below on the left—the centre first moves inertially towards the wall, then slows down because reflection is a gradual process, and finally moves back towards the right with approximately the same speed. The fluctuations in velocity are due to interference among the reflected waves.

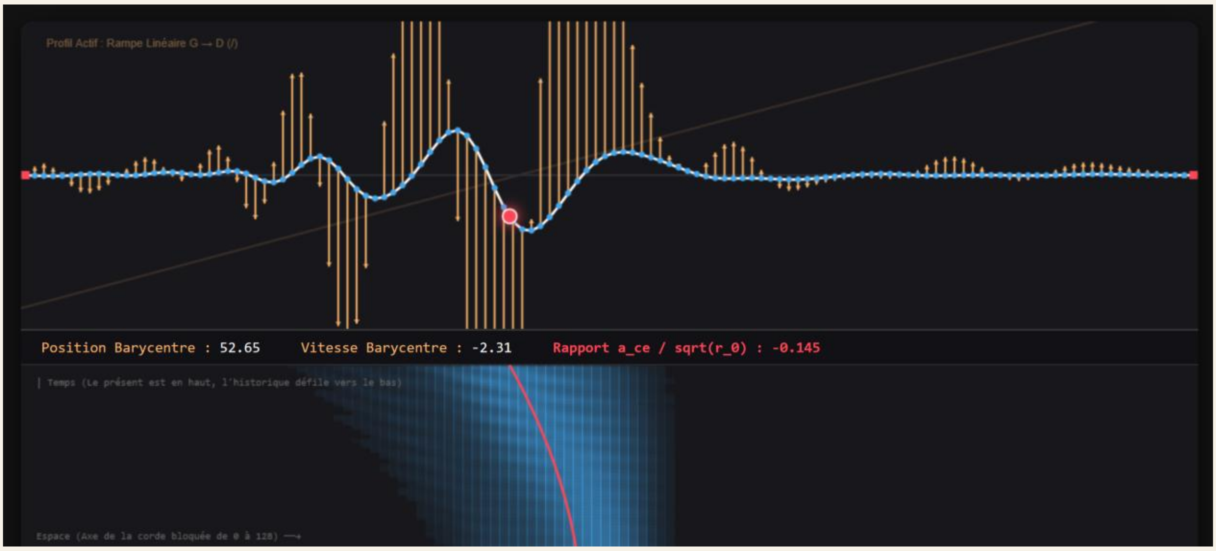






b. **Acceleration of the Centre of Energy Opposite to the Gradient of a Linear Potential**

If the system is placed in a linear potential, the centre of energy accelerates in the direction opposite to the potential gradient, echoing Newton's second law: a force produces an acceleration.

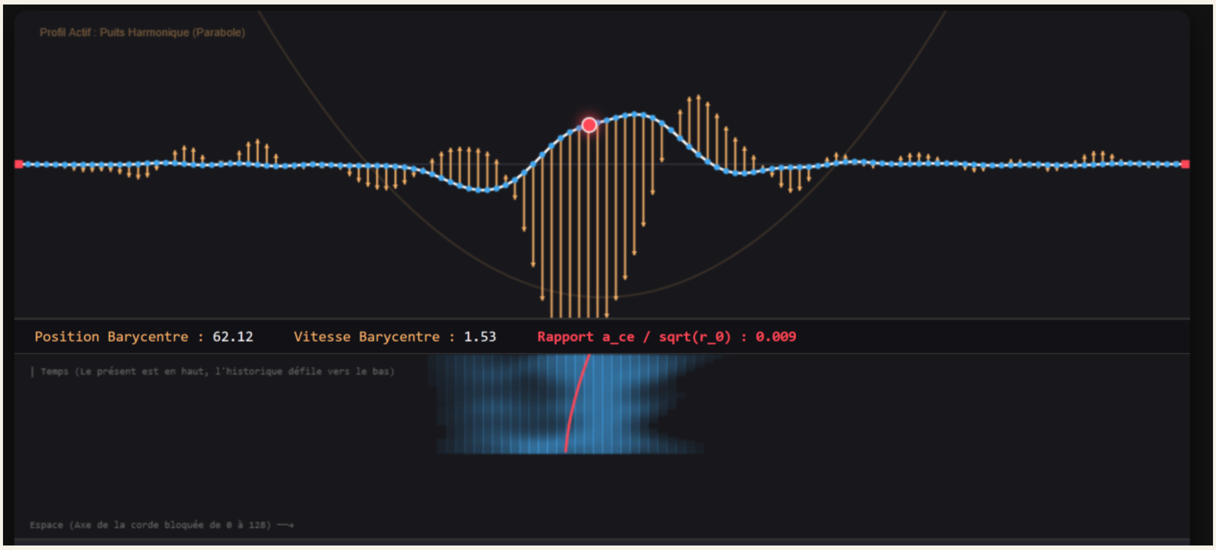


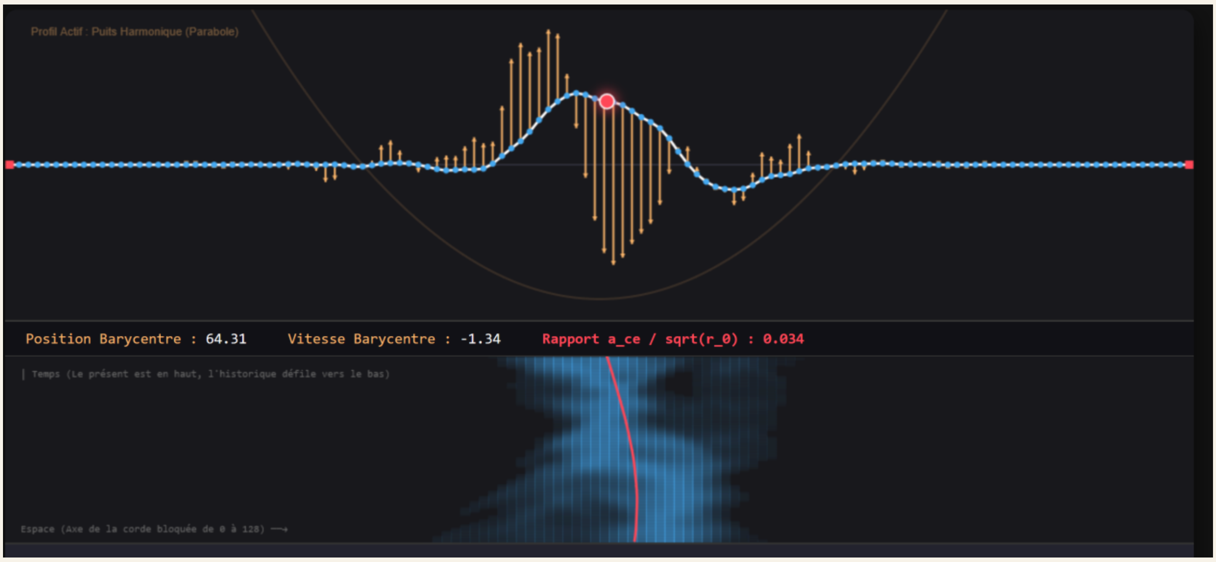
c. A Wave Packet in a Harmonic Potential

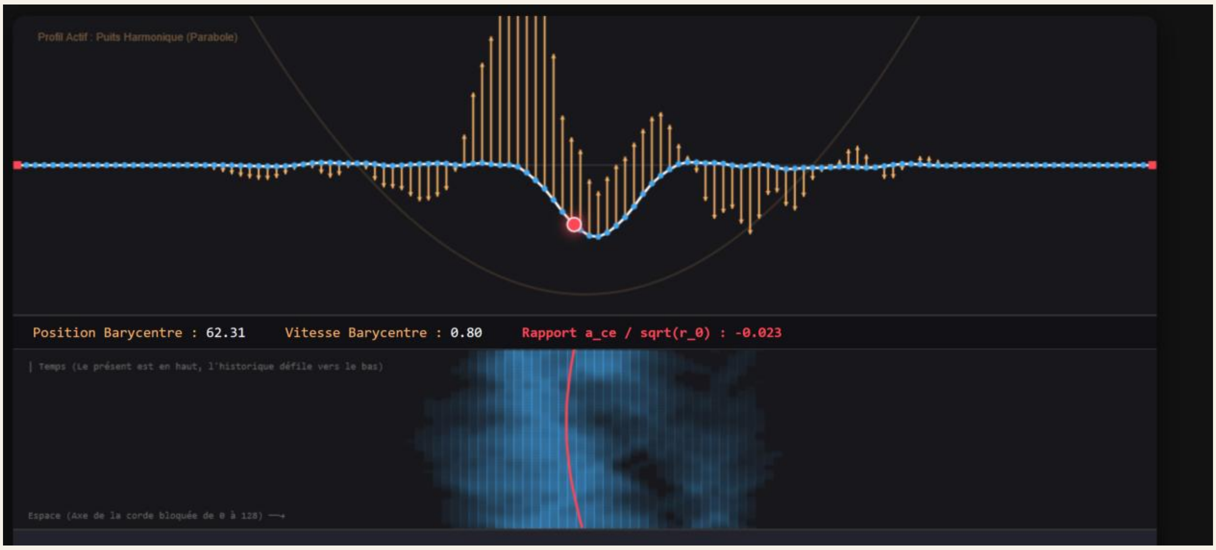
We shall now choose an additional spring stiffness whose spatial profile has the form of a harmonic potential—that is, it is proportional to the square of the distance from the centre.

In this case, the centre of energy of the wave packet behaves like… the mass of a small spring oscillating around its equilibrium position!

(see Interactive Animation 11)







Naturally, under the deliberately simple conditions considered here, Newtonian laws appear only very approximately, and only in the behaviour of the packet's centre of energy. Later we shall identify the specific conditions under which Newtonian dynamics emerges and examine regimes in which the approximation is far more accurate than in the examples just considered. But the central idea is already visible: the centre of the wave packet behaves approximately like a harmonic oscillator. A set of coupled first-order harmonic oscillators, endowed with an additional stiffness profile shaped like a harmonic potential, gives rise to a second-order harmonic oscillator in its centre of energy. One may then wonder whether such second-order oscillators could themselves be coupled so as to produce third-order oscillators—or whether the supposedly first-order oscillators were never truly fundamental in the first place.

Something like this occurs in real physics. Near an energy minimum, every sufficiently smooth potential is locally approximately harmonic: the quadratic term is the first non-trivial term in its Taylor expansion. Under this approximation, an electron bound in an atom behaves like a harmonic oscillator; an atom within a molecule behaves approximately like one; and a molecule in a crystal lattice behaves approximately like one as well. At the most fundamental level, elementary particles such as electrons are themselves excitations of systems of coupled harmonic oscillators—namely, quantum fields.

III. The Free Klein–Gordon Equation

III.1 Deriving the Free Klein–Gordon Equation

We shall now begin to express our system of coupled springs mathematically. For the moment, we shall set aside the additional stiffness profile and focus on the homogeneous chain considered earlier.

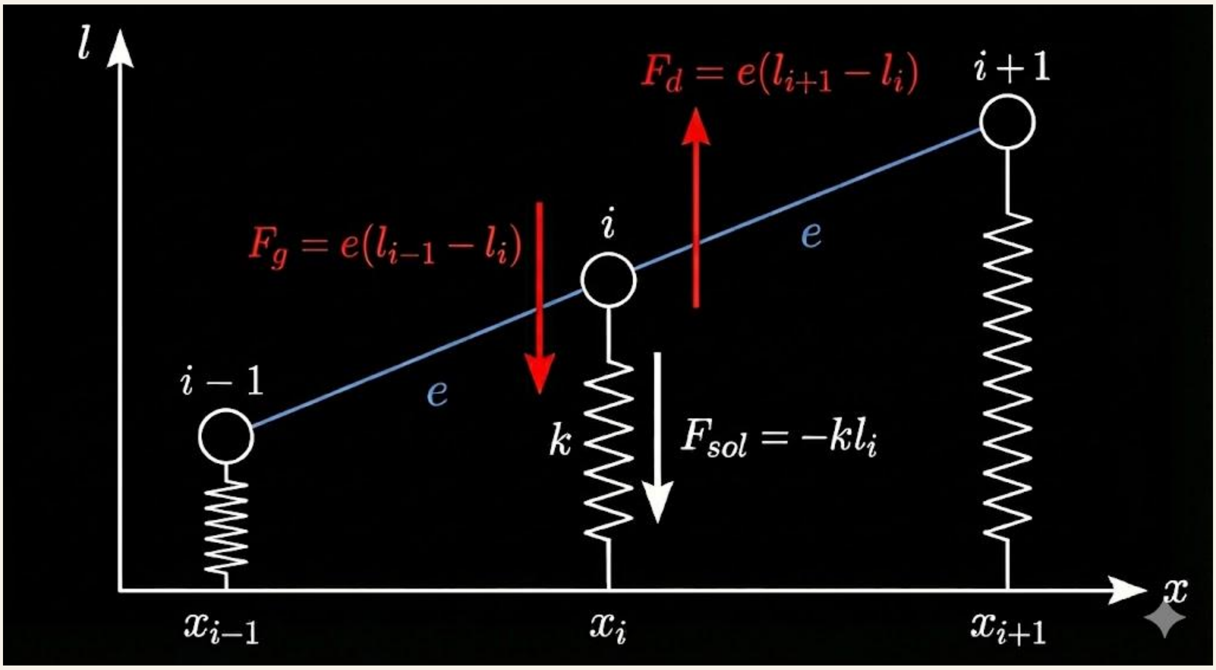


Figure: force balance on one mass

We shall focus on a single site, site i, and determine the forces acting on mass i. There is first the restoring force exerted by the local spring, whose form we already know. There are also the forces exerted by the elastic couplings connecting mass i to its two neighbours.

We have: $F_i = -rl_i + e(l_{i-1} - l_i) + e(l_{i+1} - l_i$

That is: $F_i = -rl_i + e(l_{i-1} + l_{i+1} - 2l_i)$

where κ is the elastic constant of the coupling. Strictly speaking, the vertical force exerted by the coupling is the projection of its tension onto the vertical axis, but within the small-displacement approximation used here this distinction can be ignored.

We therefore obtain: (12) $m\frac{d^2 l_i}{dt^2} = -rl_i + e(l_{i-1} + l_{i+1} - 2l_i)$

We now introduce the variable x_i = i a, where a is the distance between two successive springs. In other words, x_i is the horizontal coordinate identifying the position of a given oscillator.

$$(13)\ m\frac{\partial^2 l(x,t)}{\partial t^2} = -rl(x,t) + e\left(l(x-a,t) + l(x+a,t) - 2l(x,t)\right)$$

This is the dynamical equation of our discrete spring system. Before proceeding, note that both Figure 7 and the equation show that if the central mass lies exactly at the average height of its two neighbours, the resultant force from the elastic couplings vanishes.

The energy associated with each mass is:

$$(14) E(x,t) = \frac{m}{2}\left(\frac{\partial l(x,t)}{\partial t}\right)^2 + \frac{r}{2}l(x,t)^2 + \frac{e}{2}(l(x-a,t) - l(x,t))^2 + \frac{e}{2}(l(x+a,t) -$$

We shall now return to the dynamical equation and take the limit in which a becomes very small—that is, the continuous limit of the discrete spring system.

Two elementary mathematical results are required:

When a is small, we have:

$$l(x+a) \approx l(x) + \frac{\partial l(x)}{\partial x} a + \frac{1}{2}\frac{\partial^2 l(x)}{\partial x^2} a^2$$

and

$$l(x-a) \approx l(x) - \frac{\partial l(x)}{\partial x} a + \frac{1}{2}\frac{\partial^2 l(x)}{\partial x^2} a^2$$

Using these expansions in the discrete equation, we obtain the following field equation:

$$(15)\ m\frac{\partial^2 l(x,t)}{\partial t^2} = -rl(x,t) + ea^2\frac{\partial^2 l(x)}{\partial x^2}$$

This is a one-dimensional field equation. By definition, a field is a physical system described by a quantity defined at every point in space and at every instant in time. Here, the quantity is the continuous displacement ϕ(x,t), which replaces the individual displacements of the discrete masses.

We may likewise define an energy density for the model by starting from the discrete energy expression and taking the same continuous limit. This gives:

(16) $\rho(x,t) = \frac{m}{2}(\frac{\partial l(x,t)}{\partial t})^2 + \frac{r}{2}l(x,t)^2 + \frac{ea}{2}(\frac{\partial l(x,t)}{\partial x})^2$

We shall now move somewhat beyond the literal mechanical model of springs and elastic couplings and regard this equation as describing a genuine one-dimensional physical field. To make contact with relativistic notation, we introduce the parameters c and Ω through the appropriate identifications.

$$l(x,t) \longrightarrow \phi(x,t)$$

$$a\sqrt{\frac{e}{m}} \longrightarrow c$$

$$\sqrt{\frac{r}{m}} \longrightarrow \Omega$$

After a few rearrangements, the equation becomes:

$$(17) \qquad \frac{\partial^2 \phi(x,t)}{\partial t^2} = c^2\frac{\partial^2 \phi(x,t)}{\partial x^2} - \Omega^2\phi(x,t)$$

This is the one-dimensional Klein–Gordon equation. It can also be obtained by applying the quantum substitutions E → iℏ∂/∂t and p → −iℏ∂/∂x to the relativistic energy–momentum relation $E^2 = p^2c^2 + m^2c^4$. Here, however, it has emerged directly from a purely classical network of coupled oscillators.

That point deserves emphasis. We have derived the equation from a spring lattice that is neither quantum nor relativistic. Yet its mathematical structure is the same. This is not an accident: the equation expresses a very general pattern of local restoring dynamics combined with spatial coupling.

The equation contains three main terms, and each must be understood correctly. The mechanical analogy makes their roles particularly transparent.

- The first term, $\partial^2\phi(x,t)/\partial t^2$, describes the local temporal evolution of the field. It is the field analogue of acceleration: the rate of change of the rate of change of ϕ at position x.

- The second term, $c^2\partial^2\phi(x,t)/\partial x^2$, represents the spatial curvature of the field. It links the dynamics at one point to the field values in its neighbourhood and corresponds, in the mechanical analogy, to the coupling forces between neighbouring oscillators.
- The final term, $-\Omega^2\phi(x,t)$, acts as a restoring force: whenever the field departs from zero, it tends to drive it back towards that value.

  The Klein–Gordon equation is therefore the direct continuous-field counterpart of the discrete equation governing our chain of coupled harmonic oscillators.

acceleration of a mass = elastic-coupling force + spring force

III.2 Energy Density of the Free Klein–Gordon Field

The corresponding energy density is:

$$\rho(x,t) = \frac{1}{2}\left[\frac{1}{c^2}\left(\frac{\partial\phi}{\partial t}\right)^2 + \left(\frac{\partial\phi}{\partial x}\right)^2 + \Omega^2\phi^2\right] \tag{18}$$

This expression is not obtained by a naive one-to-one replacement of every quantity in the discrete mechanical energy. The continuous limit changes how the coupling energy is distributed in space, and the field expression must be constructed so that the total energy is conserved.

Let us now examine the three contributions to the energy density:

- $\frac{1}{2}(1/c^2)(\partial\phi/\partial t)^2$ is the kinetic-energy contribution in the mechanical analogy; for the field, it measures the energy associated with temporal variation.
- $\frac{1}{2}(\partial\phi/\partial x)^2$ is the potential-energy contribution associated with spatial coupling. It may be interpreted as gradient or curvature energy, since it grows when the field varies from one point to another.
- $\frac{1}{2}\Omega^2\phi^2$ is the restoring-potential contribution. It expresses the energetic cost of giving the field a non-zero value.

III.3 Solutions of the Free Klein-Gordon Field

We must now discuss the solutions of this equation. In doing so, we shall introduce a physical and mathematical concept of fundamental importance in contemporary physics. It is difficult to overstate its significance: the same structure will reappear throughout the rest of this series.

We saw earlier that the equation of motion of a harmonic oscillator has sinusoidal solutions in time. We also saw that, in order to obtain permanent regimes in a spatially extended system, the spatial profile must itself be sinusoidal. The solutions of the free Klein-Gordon equation therefore combine a sinusoidal dependence on space with a sinusoidal dependence on time.

a) Right-Moving Travelling Mode

This is a sinusoidal wave travelling from left to right at a speed v. Its general form is:

$$\phi_R(x,t) = A\ \cos(\varphi(x,t))$$

Initial profile. At time t = 0, the wave has a sinusoidal spatial profile characterised by the wavenumber k:

$$\phi_R(x,0) = A\ \cos(kx + \varphi_0)$$

Propagation to the right. Saying that the wave travels to the right at speed v means that, at time t, its profile is identical to the profile at t = 0, but translated a distance vt to the right. The value observed at position x and time t is therefore the value that was located at x - vt at the initial time:

$$\Phi_D(x,t) = \Phi_D(x - vt, 0) = A\cos(k(x - vt) + \varphi_0) = A\cos(kx - kvt + \varphi_0))$$

Identification. At a fixed point in space, the field must oscillate with angular frequency ω. The expression above shows that, for fixed x, the observed angular frequency is kv. We must therefore have:

$$kv = \omega \ \text{ hence } v = \frac{\omega}{k}$$

This speed, at which surfaces of constant phase propagate, is called the phase velocity. The wave can therefore be written in its final form as:

$$\phi_R(x,t) = A\cos(kx - \omega t + \varphi_0)$$

b) Left-Moving Travelling Mode

The reasoning is exactly the same, except that the translation now takes place towards the left. At time t, the value observed at position x is the value that was initially located at x + vt:

$$\Phi_G(x,t) = \Phi_G(x + vt, 0) = A\cos(k(x + vt) + \varphi_0) = A\cos(kx + kvt + \varphi_0))$$

Identifying the angular frequency at fixed x once again gives kv = ω, and we obtain:

$$\phi_L(x,t) = A\cos(kx + \omega t + \varphi_0)$$

The only difference from the right-moving travelling mode is therefore the sign of the term containing t: a plus sign corresponds to propagation towards the left, while a minus sign corresponds to propagation towards the right.

c) Cosine Standing Mode

For this mode, the spatial profile always has the same cosine form:

$$\phi_C(x) = A\ \cos(kx)$$

Each point nevertheless oscillates sinusoidally in time with angular frequency ω. The complete cosine standing wave is therefore:

$$\phi_C(x) = A\ \cos(kx)\cos\,(\omega t + \varphi_0)$$

d) Sine Standing Mode

Exactly the same principle applies here, except that the spatial profile is a sine. We therefore obtain:

$$\phi_S(x) = A\ \sin(kx)\ sin\,(\omega t + \varphi_0)$$

e) Linearity of the Equation

Just as the equation of a harmonic oscillator is linear - meaning that if ϕ_1 and ϕ_2 are any two solutions, then every linear combination $\alpha\phi_1 + \beta\phi_2$ is also a solution - the Klein-Gordon equation is linear as well. Consequently, for example:

$$\alpha\phi_C(x) + \beta\phi_S(x)$$

is also a solution of the equation.

III.4 The Vector Space of Solutions

We have already seen that these solutions stand in relations to one another that closely resemble the relations among basis vectors belonging to two different orthonormal bases. We shall now make this analogy precise.

A. Normalising the Solutions

To define the 'size'—that is, the norm—of our waves, we use an inner product obtained by averaging their product over one period. We shall define the inner product of two solutions accordingly:

$$\langle f, g\rangle = \frac{1}{T \cdot \lambda} \int_0^T \int_0^\lambda f(x,t) g(x,t)\, \mathrm{d}x \mathrm{d}t$$

The norm of a solution is then given by:

$$\sqrt{\langle f, f\rangle} = \sqrt{\frac{1}{T \cdot \lambda} \int_0^T \int_0^\lambda f(x,t) f(x,t)\, \mathrm{d}x \mathrm{d}t}$$

In other words, the norm of a solution is simply the square root of its mean square value—that is, the square root of the average squared departure from zero.

Rather than carrying out lengthy integral calculations to determine the norm of a travelling wave, we can use a simple trigonometric identity.

- For any angle, $\cos^2\theta + \sin^2\theta = 1$. The average of their sum over a complete period is therefore also equal to 1.
- By symmetry, over one complete period, a squared cosine spends exactly as much time above and below its mean as a squared sine. Their averages are therefore identical.

It follows immediately that the average of either $\cos^2\theta$ or $\sin^2\theta$ is one half.

- For travelling waves, the squared norm is therefore proportional to one half. To obtain a unit norm, we divide by the norm—or equivalently multiply the wave by $\sqrt{2}$. The normalised travelling waves are thus:

$$e_d = \sqrt{2} \cos(\omega t - kx) \quad \text{et} \quad e_g = \sqrt{2} \cos(\omega t + kx)$$

We can verify that their inner product vanishes, as it should for two orthogonal vectors.

- For standing waves, the signal is the product of two independent sinusoidal functions, one in space and one in time. Its mean square value is therefore the product of the two mean square values, namely 1/2 × 1/2 = 1/4. To normalise it to unity, we must multiply it by 2. The normalised standing waves are therefore:

$$e_c = 2 \cos(\omega t) \cos(kx) \quad \text{et} \quad e_s = 2 \sin(\omega t) \sin(kx)$$

We can again verify that their inner product vanishes, confirming that these two solutions correspond to orthogonal vectors.

B. The Trigonometric Bridge

Let us expand the normalised right-moving travelling wave using the trigonometric identity

:

$$e_d = \sqrt{2}\ \left(\cos(\omega t) \cos(kx) + \sin(\omega t) \sin(kx)\right)$$

This is where the structure becomes transparent. We recognise our standing-wave solutions, because the two terms are precisely the cosine and sine standing modes. Substituting them gives:

$$e_d = \sqrt{2}\left(\frac{1}{2}e_c + \frac{1}{2}e_s\right) = \frac{1}{\sqrt{2}}e_c + \frac{1}{\sqrt{2}}e_s$$

Applying the same reasoning to the left-moving wave gives:

$$e_g = \frac{1}{\sqrt{2}}e_c - \frac{1}{\sqrt{2}}e_s$$

To connect this directly with the geometry of a vector space, we need only recall the elementary identity $\cos(\pi/4) = \sin(\pi/4) = 1/\sqrt{2}$. The equations immediately become:

$$e_d = \cos\left(\frac{\pi}{4}\right)e_c + \sin\left(\frac{\pi}{4}\right)e_s$$
$$e_g = \cos\left(\frac{\pi}{4}\right)e_c - \sin\left(\frac{\pi}{4}\right)e_s$$

C. Geometric Interpretation

Imagine an ordinary two-dimensional coordinate system in which the horizontal axis is defined by the cosine standing-wave vector and the vertical axis by the sine standing-wave vector. These two vectors are perpendicular and have unit norm, so they form an orthonormal basis.

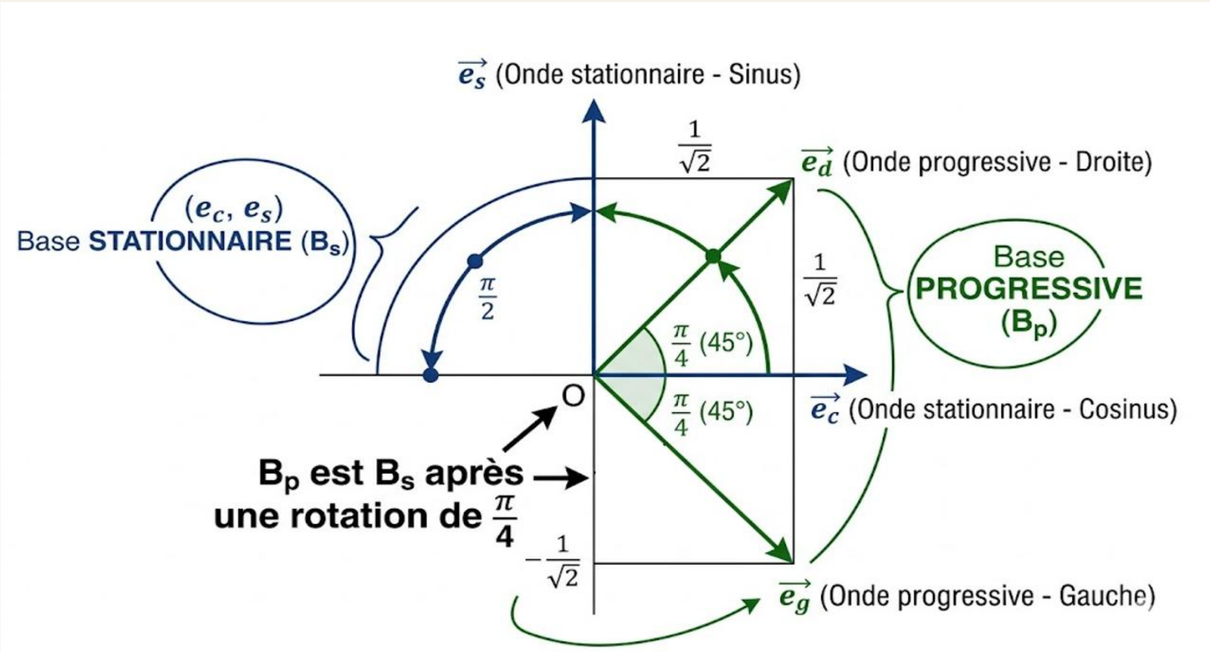
In this coordinate system, the normalised right-moving travelling wave has coordinates $(1/\sqrt{2}, 1/\sqrt{2})$. This is exactly a unit vector lying on the angle bisector, at 45° from each axis.

Likewise, the normalised left-moving travelling wave has coordinates $(1/\sqrt{2}, -1/\sqrt{2})$. It also lies at 45° from the cosine standing-wave axis, but in the opposite direction.

If we consider the figure formed by these four vectors:

- the right-moving and left-moving travelling-wave vectors are separated by an angle of $\pi/2$. They are therefore orthogonal and form a second orthonormal basis.
- The basis defined by the travelling waves is simply the basis defined by the standing waves after a global rotation of $\pi/4$.

The physics therefore shows us that the space of waves is indifferent to the way we choose to describe it. We may use travelling waves or standing waves; both are merely different coordinate systems on the same underlying space of solutions.



This vector representation of the solutions is a direct expression of the fundamental linearity of the Klein–Gordon equation. Every solution can be built from basis solutions by linear combination.

For example:

$$e_\phi = \alpha e_c + \beta e_s$$

is also a solution of the Klein–Gordon equation.

One final clarification is important. The two-dimensional vector space just described corresponds only to the solution space associated with one fixed angular frequency ω—that is, with a single mode. Every allowed value of ω has its own two-dimensional space. The full solution space of the field is therefore infinite-dimensional.

III.5 The Dispersion Relation

We shall now substitute an arbitrary basis solution into the Klein-Gordon equation, for example:

$$\frac{\partial^2 \phi_D(x,t)}{\partial t^2} = c^2 \frac{\partial^2 \phi_D(x,t)}{\partial x^2} - \Omega^2 \phi_D(x,t) = 0$$

This gives directly:

Eliminating the common factor yields what is called the dispersion relation of the field:

$$\boxed{(19)\quad \omega^{\mathbf{2}} = \mathbf{k^2 c^2} + \mathbf{\Omega^2}}$$

Let us draw out everything contained in this result. First, the angular frequency ω is no longer fixed solely by the constants of the system. The only requirement is that ω and k jointly satisfy equation (19). For every possible value of k - that is, for every real number - there is a corresponding value of ω. There are therefore infinitely many solutions of this type, each associated with a different value of k.

Let us now recall what a solution of this form means. If we focus on the spatial point x = 0, then the field varies sinusoidally in time, exactly as in the harmonic oscillator. We recover the familiar relation T = 2π/ω between the temporal period T and the angular frequency ω.

Now fix t and examine the solution across space. At any given instant, the field is a sinusoidal function of x. This sinusoid repeats after one complete 'spatial period', corresponding to a phase change of 2π. This spatial period is denoted by λ and is, as we have already seen, called the wavelength. We recover the relation:

$$(20)\ k = \frac{2\pi}{\lambda}$$

This is the spatial counterpart of the relation between angular frequency and temporal period.

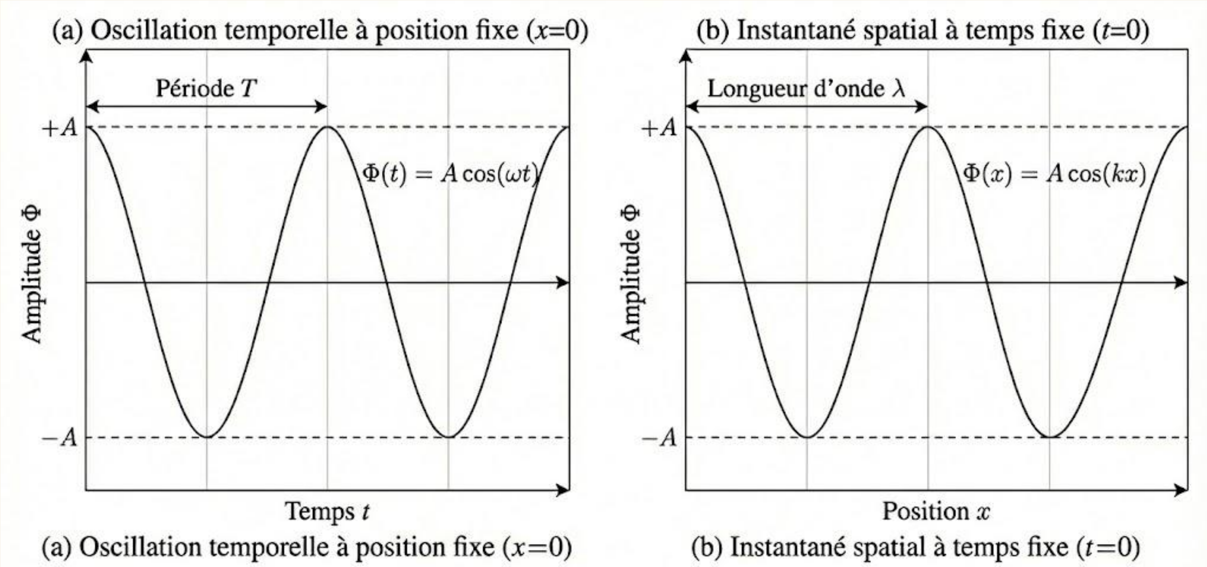


Figure 8 : Visualisation d'une solution d'onde cosinusoïdale monochromatique. (a) Évolution de l'amplitude du champ au cours du temps en un point unique de l'espace. (b) Profil spatial instantané du champ à un moment unique.

Choosing a different instant t merely changes the phase. The solution retains the same shape; its profile is simply translated through space.

We can now determine precisely how this translation occurs. Suppose that the phase is zero at x = 0 when t = 0, which we are always free to choose by fixing the arbitrary origin of phase. At a later time t, the location at which the phase is zero has moved. Combining the temporal and spatial phase relations shows that the point of constant phase has travelled a distance x = (ω/k)t. The same argument applies to any phase. Physically, the phase therefore moves at the speed:

$$(21) \quad v_\varphi = \frac{\lambda}{T} = \frac{\omega}{k}$$

This quantity is called the phase velocity. We obtained the same result earlier and have now recovered it in a different way. For positive k, the solution is therefore a sinusoid travelling to the right with speed v_phase = ω/k. In physics, such a solution is called a monochromatic travelling wave: a propagating wave with one sharply defined wavelength.

It is important to note that, for the Klein-Gordon equation, the phase velocity depends on the wavenumber. Using the dispersion relation, we find:

$$v_\varphi = \frac{\sqrt{k^2c^2 + \Omega^2}}{k} = \sqrt{c^2 + \frac{\Omega^2}{k^2}}$$

In the special case Ω = 0, however, the phase velocity no longer depends on k and is equal to c. The same constant is also the limiting value approached by the phase velocity as k tends to infinity. It therefore appears as a characteristic speed of the field.

The Klein-Gordon equation also admits a solution with infinite wavelength, corresponding to k = 0. In that case, equation (19) gives the remarkable result:

$$(22) \quad \omega_0 = \Omega$$

If you follow the argument through to the final part, we shall see the deeper meaning of this equation in quantum field theory. For the moment, its classical significance is that even when the wavelength is infinite, the field still possesses a non-zero oscillation frequency. In the mechanical analogy, this is the configuration in which all the springs oscillate in phase. This does not mean that a completely static zero field is impossible: setting the amplitude to zero gives a field that vanishes everywhere and at all times.

III.6 The Klein-Gordon Equation as a Collection of Harmonic Oscillators

We have now reached the most important point of this first article. We saw that the Klein-Gordon equation, like the equation of the harmonic oscillator, is linear. We also saw in Section II that the constant-frequency solutions of our network of coupled harmonic springs are configurations in which every spring follows the same temporal trajectory as a harmonic oscillator, but with a restoring coefficient modified by the curvature of the spatial mode.

Consider two standing-wave solutions of the equation:

$\phi_1 = \sin(k_1 x)\, sin\,(\omega_1 t)$ and $\phi_2 = \sin(k_2 x)\, sin\,(\omega_2 t)$.

The linearity of the equation implies that:

$$\phi = a\phi_1 + \ b\phi_2$$

where a and b are arbitrary numbers, is also a solution. If we denote the time-dependent amplitudes by α(t) and β(t), the two standing-wave solutions may be rewritten in separated form:

(23) $\phi_1 = \alpha(\mathrm{t}) sin\,(k_1 x)$ and $\phi_2 = \beta(\mathrm{t}) sin\,(k_2 x)$

Only the amplitudes α(t) and β(t) depend on time; the spatial profiles remain fixed. Substituting each solution separately into the Klein-Gordon equation gives:

$$\frac{\partial^2 \phi_1(x,t)}{\partial t^2} = c^2 \frac{\partial^2 \phi_1(x,t)}{\partial x^2} - \Omega^2 \phi_1(x,t)$$
$$\Longleftrightarrow (24) \qquad \frac{\partial^2 \alpha(t)}{\partial t^2} = -[\Omega^2 + {k_1}^2 c^2]\alpha(t)$$

and

$$\frac{\partial^2 \phi_2(x,t)}{\partial t^2} = c^2 \frac{\partial^2 \phi_2(x,t)}{\partial x^2} - \Omega^2 \phi_2(x,t)$$
$$\Longleftrightarrow \qquad \frac{\partial^2 \beta(t)}{\partial t^2} = -[\Omega^2 + {k_2}^2 c^2]\beta(t)$$

to be compared with the ordinary harmonic-oscillator equation:

$$(3) \quad \frac{d^2 l(t)}{dt^2} = \ -\frac{r}{m} l(t)$$

In other words, the amplitude of each standing wave obeys the equation of a harmonic oscillator whose squared natural angular frequency is $\Omega^2 + c^2k^2$. This is exactly what our qualitative study of the coupled-spring network already suggested. Here Ω^2 corresponds to the original local restoring stiffness, while c^2k^2 plays the role of the additional stiffness generated by the spatial curvature of the coupling.

If we now substitute the sum of the two standing-wave solutions into the equation, we obtain:

$$\left(\frac{d^2\alpha(\mathrm{t})}{dt^2} + \left[\Omega^2 + {k_1}^2 c^2\right]\alpha\ \ (\mathrm{t})\right)\sin(k_1 x) + \left(\frac{d^2\beta(\mathrm{t})}{dt^2} + \left[\Omega^2 + {k_2}^2 c^2\right]\beta\ \ (\mathrm{t})\right)\sin(k_2 x) = 0$$

For this equation to hold for every value of x, the two independent spatial profiles must satisfy their own equations separately:

$$\frac{\partial^2 \alpha(t)}{\partial t^2} = -\left[\Omega^2 + {k_1}^2 c^2\right] \alpha(t)$$

and

$$\frac{\partial^2 \beta(t)}{\partial t^2} = -[\Omega^2 + {k_2}^2 c^2]\beta(t)$$

Each mode therefore obeys the dynamics of a harmonic oscillator independently of every other mode. This result generalises to any superposition of standing waves: each standing mode can be described as an independent harmonic oscillator, dynamically decoupled from the others. More generally, a free field—that is, a field not coupled to another physical system—whose dynamics are linear can be regarded as an ensemble of dynamically independent harmonic oscillators.

Together with the principle of combining—or, as we shall increasingly say, superposing—solutions, this is one of the most powerful ideas in contemporary physics.

It is important to begin thinking of the field as a superposition of its modes. This viewpoint is useful in classical field theory, but it becomes absolutely indispensable in quantum field theory.

The same idea is especially illuminating from an energetic point of view. Using expression (18), substituting a single mode into the energy density, and integrating over all space, one finds that each mode n carries an energy:

$$(25) \qquad E_n \propto A^2 \omega_n^2$$

The same result can be obtained in another way. Starting from the mode equation and recalling that the energy of a harmonic oscillator is given by expression (11), one arrives at exactly the same formula.

The energy stored in each mode therefore depends not only on the square of its amplitude, as for any harmonic oscillator, but also on the square of its own natural angular frequency. The two descriptions are perfectly consistent.

III.7 Group Velocity

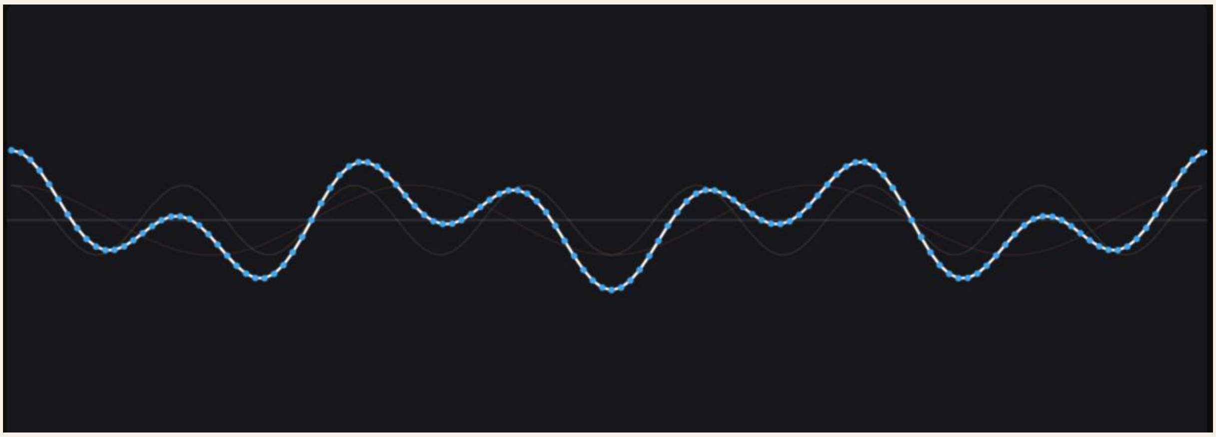
We shall now take a combination of several travelling waves and analyse how they behave.

We shall begin with the simplest possible case by combining two solutions.

$\phi_1 = \mathrm{A}_1 \cos(k_1 x - \omega_1 t)$ and $\phi_2 = \mathrm{A}_2 \cos(k_2 x - \omega_2 t)$

We therefore have:

$$\phi = A_1\, cos(k_1 x - \omega_1 t) + A_2\, cos(k_2 x - \omega_2 t)$$



At t = 0 and x = 0, these two solutions have the same phase. They therefore add constructively at that time and place, where both are at their maximum. Elsewhere, because their wavenumbers differ, they are out of phase and no longer add in the same way. At points where they are in opposite phase, the two solutions interfere destructively and may partially or completely cancel, depending on their respective amplitudes. This spatial pattern of addition, subtraction, and every intermediate case is what physicists call interference.

We shall focus on the regions of large amplitude - the envelope peaks. At what speed do these regions move? One might be tempted to answer: at the phase velocity. But two monochromatic waves with different wavelengths do not, in general, have the same phase velocity. The answer must therefore be subtler.

To see this, let us define:

$$k_2 = k_1 + \Delta k$$

$$\omega_2 = \omega_1 + \Delta\omega$$

The condition for the two waves to have the same phase is:

$$k_1 x - \omega_1 t = (k_1 + \Delta k)x - (\omega_1 + \Delta\omega)t$$

That is:

and therefore:

$$\Delta k x - \Delta\omega t = 0$$

The velocity of this equal-phase point is thus:

$$(26) \qquad v_g = \frac{x}{t} = \frac{\Delta\omega}{\Delta k}$$

We shall now generalise the result.

Suppose that, instead of only two waves, we have a continuous superposition of travelling waves:

$$\phi = \int A_k \cos\,(kx - \omega(k)t)dk$$

At a given point, the waves must have approximately the same phase if they are not to cancel one another. To make this idea concrete, consider three waves whose wavenumbers are very close. For them to interfere constructively at position x and time t, their phases must be nearly equal.

In other words, the phase must vary as little as possible when k changes. Ideally, for the region in which the waves remain synchronised to persist, this first-order variation should vanish.

Formally, this condition is expressed as:

$$(27) \quad \frac{d\varphi}{dk} = 0$$

But the phase is:

The condition therefore becomes:

$$x - \frac{d\omega}{dk}t = 0$$

Hence:

$$(28) \qquad \frac{d\omega}{dk} = \frac{x}{t} = v_g$$

This expression gives the speed at which the equal-phase region - and therefore the coherent concentration of energy - moves. The quantity dω/dk is called the group velocity. It is the velocity of a wave packet: a superposition of waves that are initially synchronised within a limited region.

For many fields, however, the velocity of each monochromatic component depends on k, so the packet tends to disperse over time. The group velocity also depends on the packet's central wavelength. In general, a well-localised packet (small spatial width Δx) requires a comparatively broad spread of wavenumbers (large Δk) around the central value, and conversely. The resemblance to the uncertainty principle named after a famous German physicist is not accidental.

For the Klein-Gordon equation, combining equations (19) and (28) gives:

$$(29) \quad v_g = \frac{d\omega}{dk} = \frac{kc^2}{\sqrt{k^2c^2 + \Omega^2}}$$

When Ω = 0, or as k tends to infinity, the group velocity approaches c. Interesting...

We have not yet specified the numerical value of c. For the Higgs field - the only known fundamental field described by a scalar field of the Klein-Gordon type - c is, of course, the speed of light. It might more revealingly be called the characteristic propagation speed of fundamental fields.

Conclusion

Over the course of this chapter, we have gradually reconstructed the concept of a field from a simple but physically rich mechanical model: a chain of coupled oscillators. This detour through analogy was not a gratuitous pedagogical device. It allowed us to see concretely how local dynamics, based on interactions between neighbouring degrees of freedom, can give rise to a continuous description in terms of a field.

By taking the continuous limit of the discrete system, we obtained a genuine field equation whose structure is that of the Klein-Gordon equation. Far from appearing as an abstract construction descending from nowhere, the equation can be understood as the natural generalisation of a network of coupled oscillators: the temporal term describes the local evolution of the field, the spatial term encodes the coupling between neighbouring points, and the final term plays the role of an intrinsic restoring force.

We also introduced an idea that is absolutely central to all of field physics: decomposition into normal modes. Each mode behaves like an independent harmonic oscillator, and every configuration of the field can be regarded as a superposition of such modes. Because the dynamics are linear, interference, wave packets, and localised concentrations of energy moving through space can emerge from these superpositions.

Finally, our study of mode superposition led us to the concept of group velocity and to the emergence of wave packets whose energy propagates coherently. These packets provide a decisive conceptual bridge. They are not fundamental particles, but for a limited time they can reproduce particle-like dynamics to a remarkable degree.

Transition to Part 3: The Dynamics of Particles as Classical Wave Packets

In the next chapter, we shall take an essential conceptual step. We shall leave behind the study of idealised monochromatic modes and concentrate on localised superpositions: wave packets. We shall see that, even within a purely classical field-theoretic framework, such packets already possess many of the properties ordinarily associated with particles: spatial localisation, propagation at a well-defined velocity, effective interaction with a potential, and energy transfer.

Most strikingly, under suitable conditions the centre of energy of a wave packet can follow dynamics close to those of a Newtonian particle. We shall examine in detail the conditions under which this quasi-particle behaviour is valid, together with its fundamental limitations, most notably the unavoidable dispersion of wave packets. We shall also see that several dynamical properties commonly associated with relativistic particles are already encoded in the very structure of relativistic field equations. Finally, we shall show that a number of phenomena often presented as specifically quantum have classical analogues in field theory: interactions between quasi-particles mediated by a field, structures analogous to quasi-antiparticles, effective annihilation through interference, and the appearance of an effective mass in contexts involving spontaneous symmetry breaking.

This analysis will prepare the way directly for the central idea of quantum field theory: particles no longer appear as fundamental objects, but as localised, quantised excitations of fields. Their classical counterparts - wave packets - already provide a first and deeply illuminating conceptual approximation.