# When does a wave arrive?

Cambridge Ultrasonics Ltd

October 2016 – May 2017

David R Andrews

Having worked in R&D consulting on ultrasound projects for many years, a common discussion with clients has been, "When does a wave arrive?"

The subject often arises in discussions with clients with products in the following fields: oil and gas metering, water metering, domestic gas metering, level measurement, sports instruments, ultrasonic communications, medical imaging, imaging in concrete (and other materials) and non-destructive testing (NDT).

This is a semi-whimsical discussion of the problems associated with sorting out "When does a wave arrive?"

The discussion will be split into more than one part and it will be appear in full in the White Paper section of www.cambridgeultrasonics.com web-site in the future. The web-site also shows some video of ultrasonic pulses travelling.

## Contents



## Part 1 – Arrival Time?

Firstly, let's talk about trains.

When exactly does a train arrive at a station if it doesn't stop? Is it when the very front of the train reaches the beginning of the platform? Is it when the very last of the train leaves the last of the platform? Is it when the end of the third carriage is level with the pedestrian bridge crossing the platforms? All three suggestions are arbitrary; the last seems bizarrely arbitrary but two out of three are used commercially.

When the very front of the train reaches the first part of the platform – that is a popular choice in many applications using ultrasound, particularly NDT – but it has its problems.

Personally, I like the choice of the middle of the train being level with the middle of the platform - it's like the arrival of maximum energy and it's a robust method.

Trains and bursts of ultrasonic waves are both of finite duration, or are they? Within them there is a periodic structure (wavelength = train carriage length, with frequency = how many carriages pass a stationary point in one second). What has frequency and wavelength got to do with when a wave arrives?

Maybe a few extreme examples will help.

When does a continuous wave of constant frequency arrive? Think of an infinitely long train. It's made of equal sized carriages. When does it arrive at a station (assuming the station is not a terminus – not that a terminus would be any use for off-loading passengers from such a long train)? You turn up at the station and the infinitely long train is already passing through; you leave the station an hour later and the infinitely long train is still passing through – in fact you can leave the station whenever you want and the same train is still passing through. Bizarre example, I know, but important. When did that infinitely long train arrive? The answer is - we don't know; there is infinitely long uncertainty in the time. Just keep in mind that for a continuous wave or an infinitely long train we cannot tell when it arrives and it has an infinitesimally narrow range of frequencies (carriages per second) because they are all of the same length. In communications, an infinitely long wave of constant frequency carries no information. Agreed, if some of the carriages were a different length then you could spot a change as the train goes by and use that to convey information – like when the train arrives – but when all the carriages are the same there is no information. This long train has a lot of energy – you cannot miss it – some design engineers are enthusiastic about getting more energy out of their transducer to improve pulse detection but then end up with long trains and problems about when the pulse arrives (NDT of concrete).

Let's consider the opposite extreme. When does an infinitesimally narrow pulse arrive? Think now of the smallest train you can possibly imagine, shorter than one human being wide – another bizarre train. Think of a train as thin as a knife blade travelling through a station that is also only one knife blade's width long. This bizarre example is handy to consider (and one that many clients try to construct in their equipment). Definitely no problem measuring when the knife-train arrives: it's when the two knife-blades line-up and since the train doesn't stop at the station the arrival event will be very short indeed – ideal for accurate timing. This is the nirvana of ultrasound – engineers like short ultrasonic pulses for timing measurement. Now, there is almost no uncertainty in when the train arrives. What about the number of carriages per second? We don't know – there is infinite uncertainty in the frequency – this tells us it's going to be difficult to construct this train when we don't know the length of the carriages. How much energy does this narrow knife-blade train have? Answer – very little and that is a problem. Just keep in mind a narrow pulse and a train that is only a knife-edge long both have an infinitely wide range of frequencies but we can know very precisely when they arrive.

Virtually all waves in materials suffer distortion as they travel, the same is true of ultrasound. Distortion of waves is caused by: dispersion, non-linear propagation, diffraction and random scattering.

The next episode will explore the effect of trains as they become more like real ultrasonic pulses by travelling through real materials.

## Part 2 – Dispersion

Let’s talk trains again.

Imagine a train that starts off as a thin, tall knife blade. As it travels it gets distorted and its height gets smaller and smaller but its length gets longer and longer (it has to be this way to conserve energy). It doesn’t get any thinner so it gets longer and as it gets longer new carriages appear in it (this is definitely a weird train) but now the carriages have a range of lengths and the lengths keep changing, generally always getting longer. When does this wave arrive? Let’s assume we still have a station that is a knife blade’s width long but when do we judge that the train has arrived if the train has been stretched-out and is continuing to be stretched more and more? This is precisely the situation that many applications of ultrasound face – what a nightmare for measurement!

Just think about the dilemma you are faced with: if you measure when the leading edge of the first carriage arrives at the start of the station it will be earlier than when the last bit of the last carriage leaves – so you get a range of different times of arrival when you expected only one (you were expecting a knife-width long train). This effect is known as dispersion and it is caused by the material through which the ultrasonic waves travel; for the train analogy, the carriages of this train do not all travel at the same speed and it’s got something to do with the tracks – weird again. Well, you could measure when the first part of the train arrives (fastest phase) and when the last part arrives (slowest phase) and take an average. This is getting closer to working out when the energy of the train arrives, around the middle of the train; it is possible to measure where the peak energy of the train or ultrasonic pulse arrives. Measurements of the start and end of the train determine different phase (wavelength) velocities. The fastest phase is, at first impression, the leading part of the pulse (start of the train) and it is often the lowest frequency in the pulse; the slowest phase is, at first impression, the last part of the wave (end of the train) but confusion comes from how the pulse was first created.

I like the idea of measuring when the maximum energy in the train or ultrasonic pulse arrives because with carriages changing length it’s not safe to use carriage length to measure time of arrival of the train. A nice way to measure when maximum energy arrives is if you know the shape the train will have at a particular station - then you can use correlation or matched filtering (both signal processing algorithms) to detect when the train arrives. Correlation gives you a sharp spike when peak energy arrives (the peak is the auto-correlation of the pulse).

What causes trains to change shape or ultrasonic pulses to change shape? It's called dispersion but what is the mechanism of dispersion? Firstly, any pulse of wave (of any kind) has to be constructed of a range of frequencies – this is a train with carriages of different lengths – let's put this down to being weird again. As a wave passes through a material it successively compresses then de-compresses the material so the molecules get pushed together then pulled apart. Materials with strong bonds between molecules, when the molecules are packed close together (say, in a metallic solid), move very little, particularly if each molecule is heavy – these materials behave elastically unless the wave pressure is very high. Ultrasonic waves pass through solid metals with little dispersion – pure aluminium is good example. By contrast, materials with weaker bonds where there is space around the molecule, can move much greater distances, for example polymers with long chain molecules. If the molecules start flexing or vibrating (like a skipping rope) then they can absorb energy from the ultrasound and turn it into heat. Generally, these molecules are more likely to absorb waves having wavelengths closer to the lengths of the molecules that are free to move – in polymers, the chain lengths that are relatively free to move are all microscopic, which means short wavelengths (higher ultrasonic frequencies) are absorbed more quickly. It also means the different frequencies travel at different speeds. These two effects: varying absorption with frequency and varying speed of sound with frequency, are the hallmarks of dispersion.

Remember that all ultrasonic pulses are made up of a collection of wavelengths (and frequencies). If a short pulse is created and sent through a dispersive material then the different frequencies will be differentially absorbed and they will travel at different speeds. As the pulse travels through the material, it becomes progressively distorted nearly always getting longer and less intense. Also the longer wavelengths usually travel faster than the shorter wavelengths. The very short train mentioned earlier started as a knife-width long and very high but as it passes along a dispersive track it becomes progressively more and more like that other bizarre train from Part 1 - the train that was infinitely long. Remember how difficult it was to determine when the infinitely long train arrives? You cannot make a judgement – it's impossible to judge. As the amplitude decreases it gets more difficult to measure the fastest phase (first carriage – the first carriage is shaped like a thin, pointed arrow-head) because the arrow-head is so sharp it's difficult to see it when arrives; ditto for the last carriage (slowest phase). The shape of the waves inside the pulse keeps changing too. This is a case when making measurements based upon the peak energy of arrival becomes more robust than first or last time of arrival.

Incidentally, thin plates of material, even aluminium are dispersive just because they are thin; likewise capillary waves in liquids (rather than gravity waves) are dispersive; and so too are waves running along surfaces.

## Part 3 – Bizarre in Oil and Gas

If you have not read Parts 1 and 2 of this series then jumping straight into Part 3 might leave you flummoxed and confused but the important point to understand is that I am trying to make an analogy between a pulse of ultrasonic waves and a train on a track.

Want to know about a seemingly bizarre criterion for judging when a wave arrives?

When in doubt I talk about trains, so here goes – criterion for when a wave arrives: wait until the third carriage in the train passes through the station (without stopping as usual) and then judge the train to have arrived when the end of the third carriage passes the footbridge over the tracks at the station. Why the third carriage? Why not the first or second or tenth? Why wait until the end of the carriage? Is this bizarre or what?

This seemingly bizarre method comes from the oil and gas business where they use flow-meters to measure how much oil or gas is flowing per second down a pipeline. Governments tax the quantity of oil or gas produced in a day or a week or in some chosen period of time; they require an accuracy of +/-0.1% on the volume produced. A popular type of flow-meter uses ultrasonic pulses to measure flow-speed. It typically measures the transit times for pulses to travel in opposite directions at an angle of about $45^o$ along the length of the pipeline: ultrasonic pulses travel faster when travelling with the flow (shorter transit time) and slower against the flow (longer transit time). The two transit times are subtracted, from which the speed of the flow can be calculated. Multiplying the flow-speed by the cross-sectional area of the pipe gives the volume of oil or gas that is flowing per second. It's important to know transit times accurately to achieve +/-0.1% accuracy in volume/second measurement. Unfortunately, the difference in transit times occurs in the denominator of the formula for flow-speed and that means errors can become high at low flow-speeds when the transit times become almost equal.

In oil and gas flow-meters, it is crucially important to know when a wave arrives so that transit times can be measured accurately.

Trouble is there are problems with the third-carriage method caused by the train changing shape as it travels down the track: ultrasonic waves change shape as they travel due to: dispersion, diffraction, reflection and non-linear propagation. In Part 2 of this series I discussed the effect of dispersion on ultrasonic waves … and on trains.

How does the oil and gas industry judge when the pulse arrives? This is the recurring question that is central to these articles.

Let's talk trains again.

Which carriage do they use to measure when the train arrives? Not the first: the front of the carriage is very small in height and it is quite difficult to judge when it has arrived due to noise and that makes the measurement insufficiently accurate. Not the last carriage: the end of the carriage is also very small and similarly difficult to detect accurately. Engineers generally look for the end of the second or third carriage. Which carriage is chosen is not particularly important provided the system consistently detects the end of the same carriage for every train arriving. But why choose this arbitrary carriage? The answer is: firstly, because carriages grow in height as more carriages arrive, by the third carriage the carriage-signal height has become strong and easy to detect; secondly, because the third carriage is higher, its rate of change of height with time (at the beginning and end of the carriage) is large in magnitude so the carriage-end is a sharp event to detect. Recall that the arrival of the very thin knife-width train in Part 1 could be accurately determined for the same reasons. So the third carriage-to-arrive method is not such a bizarre approach as it may at first have seemed.

Any flow-speed measurement system relying upon judging when a wave arrives requires to be calibrated because every one of them has its own, particular (arbitrary) judgement of when a wave arrives (peak energy is as arbitrary as third-carriage). In every method there is a calibration factor or fudge-factor that must be determined. The third-carriage method is no different and it too relies on accurate calibration for it to provide a flow-speed. Choose a different carriage and you have to re-calibrate the flow-meter.

Is the third-carriage flow-meter method without problems?

No.

Problem #1. Remember I mentioned it is important to be utterly consistent in finding the end of the third-carriage? Choose a different carriage by accident and you get a different transit-time – different by a whole carriage length, which is a very large error indeed.

Problem #2. In a dispersive material, carriage length is variable (see Part 2 of this article) so even if the method does not make an error in counting up to 3, variability in carriage length can generate unacceptably large errors – generally much more than 0.1%.

These two problems make the third-carriage method sensitive to pulse changes and that is the method's Achilles heel – it is not robust against pulse shape changes.

The third carriage method is <u>not</u> a peak-energy/group velocity measurement. Peak energy arrival (group velocity) can be calculated for almost any shape of the pulse so it can almost always be used to measure flow-speed. One way to summarize the differences is: third-carriage arrival measurement is a fragile method whereas peak-energy arrival is more robust.

The engineers in the oil and gas industry recognized the fragility of the third-carriage method and put sticking plaster on it to stop it from giving wildly changing flow-speed values. What should you do if you notice the speed value suddenly change? Answer: ignore the new (different) value and substitute the last speed-value you trusted. This is what they do.

Isn't this just guessing?

Yes, strictly speaking, it is guessing but there is some justification for guessing. A pipe-line has a large quantity of material in it and although pressure waves can travel along it and locally change the density these are transients and over a long period of time pressure waves can probably be ignored. Put it another way: what is the probability that a large mass of oil or gas flowing at a nominally constant speed will suddenly change its speed by more than +/-0.1%? The oil and gas industry will tell you it is small. If true then guessing that the speed has not changed is a good guess – just keep those pumps running at a constant speed.

What happens if the pressure in the pipe-line is fluctuating?

The algorithm for measuring speed is not robust against fluctuations in pressure and speed because the fundamental criterion of when the third carriage arrives is not robust. The third carriage arrival criterion is good at following a flow that is almost always constant in speed. I don't intend to go into the mathematics here but the third-carriage arrival method is so accurate it appears to break a law of physics (effectively Heisenberg's Uncertainty Principle, which considers when a wave arrives). However, the third-carriage arrival method is more complex than just Heisenberg's Uncertainty Principle because it also assumes the low probability of speed changing and guessing that the speed has not changed is not unreasonable under certain conditions. Effectively, the method is applying Baysian conditional probability to the answer.

Have I got a better method?

I've got a suggestion for a better method. Use two methods for measuring speed: the first method is the existing method, which is accurate under suitable conditions but fragile; the second method is to measure the peak energy of arrival (group velocity), which gives a result under virtually all conditions. Peak energy arrival is based upon the entire pulse and it is not tied to an arbitrarily selected carriage so peak-energy arrival should be more robust.

Peak energy arrival time may not be as accurate as the third carriage method (although under the advantageous conditions required for the third carriage method I would expect peak energy arrival to produce good accuracy too).

My suggestion is to run both methods simultaneously and cross-calibrate the two. Use the third-carriage method when there are no disturbances, for its accuracy, but when the third carriage runs off the rails then temporarily switch to using the peak energy arrival time until the third carriage method re-stabilizes.

I like to refer to this hybrid method as the Best of both Worlds. I bet those oil and gas guys are probably already using it!

Some flow-meters send pulses around the perimeter of the pipe, instead of straight across the middle of the pipe, in a swirling helical motion, to sample preferentially the flow speed close to the pipe-wall. Swirling waves when used with the conventional waves at $45^{o}$ give an indication of the flow speed profile across a diameter; the flow-speed changes across a diameter of the pipeline due to drag on the flow near the wall of the pipeline. Cambridge Ultrasonics web-site shows a video of swirling waves; the waves undergo large changes in profile but appear to return to a simple reflection of their shape once they have travelled through 180$^{o}$ of the pipe perimeter (there must be some interesting mathematics to confirm the observation). I suspect that detecting the third-carriage arrival time is even more unreliable with swirling pulses.

## Part 4 - Ultrasonic Interference and Diffraction

If you have not read Parts 1 to 3 of this series and you plan to jump straight into Part 4 then you are going to save a lot of time but bear in mind an underlying analogy between a pulse of ultrasonic waves and a train on a track.

Ultrasonic waves can be diffracted in the same way as light, which we all learn about at school. Diffraction makes use of the interference of waves: waves can add together or subtract – depending upon their relative polarity (positive displacement or negative displacement). Diffraction happens when a wave in a continuous medium meets an obstacle or discontinuity in the medium. In an ultrasonic wave, a row of molecules or atoms move in the same direction together and in doing so exert a force on the molecules in front of them, which respond by moving in the same way but just a little later. During diffraction, ultrasonic waves pass-by a region where the propagation medium is constrained in size, beyond which the medium becomes larger. Molecules at the edge of the beam, after the contrained region has been passed, exert forces on molecules off to the side of the main beam and so that these side-molecules start to move as well as molecules straight ahead (in the main beam). Movement of the molecules off to the side form a wave that has turned round the corner of the physical constraint – or diffracted.

What do these wave effects have to do with when a wave arrives?

To try and utterly confuse you I'm going to talk trains and interference/diffraction at the same time and try to bend the analogy of trains to explain what happens in ultrasonic waves; what happens in interference and diffraction is so remarkable it makes for an interesting train ride.

Remember the aim is to understand when a wave arrives.

In Part 3 I described a common method for measuring flow-speed in ultrasonic flow-meters used by the oil and gas industry. At the heart of this method is the judgement of when does a wave arrive. I gave the opinion that the method described in Part 3 is fragile because it can be easily disturbed if the shape of the ultrasonic pulse changes. In this part I'm going to consider how interference and diffraction can change the shape of an ultrasonic pulse and how that can mess-up the judgement of when a wave arrives.

Let's talk trains.

A train leaves a station. We are used to trains gradually building up speed from zero, that's because our fragile bodies cannot cope with the large inertial forces associated with high accelerations. For an ultrasonic train, however, things are totally different: the train gets started by some surface moving and the train is up to speed more or less immediately. Another odd thing about this (ultrasonic) train is that it does not stand waiting for passengers to board; it doesn't exist until it starts moving. You have to imagine some magic source out of which the train just flows with a wizard waving a wand to control the shape of the train (the wizard is the design engineer).

Here is odd fact #1: it isn't just one train that gets created, generally two trains get created and start travelling in opposite directions along the track in an ultrasonic transmitter. Imagine the station is a terminus, which is quite common in ultrasonic transmitters, and the magic source of the trains is initially a few meters from the end of the track where there are, of course, a set of buffers. Clearly, one train is going to head out of the station going at full speed; but the train going in the opposite direction hits the buffers at full speed and rebounds off them. However, it rebounds in a strange way: every bit of the train has to travel to the buffers and only when each bit reaches the buffers does it rebound. This has the effect of turning the train around. This is reflection.

This should really be another odd fact: depending upon what type of train it is (I'll come back to this) the train that initially heads to towards the buffers can be upside down (not very pleasant for the passengers!)

Odd fact #2: when the part of the train that has rebounded from the buffers meets the part of the same train that is still travelling towards the buffers then the two parts of the train pass through each other and as they do so they can add together or subtract but they cause no permanent change to the train. Eventually the whole train rebounds from the buffers unaltered by this interference. This is our first taste of the strange process of interference – the train crash that doesn't necessarily cause any lasting change. Incidentally, when the train is interfering with itself like this we observe standing waves (Cambridge Ultrasonics' web-site has a numerical simulation with video of this phenomenon).

Remember two trains were created and we have described what has happened to the one that headed towards the buffers but the other train, that set off heading out of the station, is still travelling. Only now the rebounded train is going in the same direction as the original train. Due to the perfect reflection at the buffers and the fact that interference causes no lasting change, the rebounded train is an exact copy of the original train. Actually, certain buffers can result in the reflected train being upside down on the tracks and, as previously mentioned, some buffer-trains start off upside down. For the moment, let's ignore this detail for the sake of the roller-coaster ride to come.

Odd fact #3: say the magical source of the two original trains was 5 metres from the buffers then the head of the rebounded train will be 2 x 5 m = 10 m behind the head of the train that headed directly out of the station. If both trains are longer than 10 m then they will overlap partially: (a) with the first 10 m being the original train (no overlap), (c) the last 10 m being the end of the rebound train (again, no overlap) and (b) for the middle section of this combined train, where they do overlap, they can add-up or subtract. Remember, the two trains are travelling at the same speed in the same direction so now they are synchronised – they have effectively become glued together with the interference locked in place and the train has changed shape permanently (except, whenever it is convenient we can think of the train as the addition of the two original trains or the addition of any number of arbitrary trains as long as they add up to the observed shape of the train – this is called superposition). Remember, we are interested in anything that causes a change to the shape of the train because shape can strongly affect when we judge a train to arrive – OK, here is one shape-changer already, interference. The wizard-engineer may design a transmitter-station to emit a train with a certain profile but what comes out of the station may not match what was intended because of the presence of the second, slightly delayed train!

I think we are all agreed this is a bizarre train but now the bizarreness gets worse.

Odd fact #4: as the train leaves the station the outside skin of the train behaves very strangely. It becomes another of those magical sources emitting new trains that start to head off in all possible directions – virtually all of these new trains come off the train tracks (but keep going as if new tracks were heading in all directions). This skin-source effect happens for every part of the train that passes the end of the station. What does this look like? Well, for the sake of argument let's assume the station lies in the middle of a barren plane – just to simplify the discussion. The skin-source initially forms a line around the outside perimeter of the carriages (precisely by the end of the station) and this line expands like a doughnut growing in radius as the original train continues down its track; the rate that the doughnut's radius expands is the same as the speed of the train. If it can, the doughnut will also expand back into the station.

It helps if you have just taken hallucinogenic drugs to visualize Odd Fact #4!

Right, the doughnut expands radially, with some of the doughnut (external part) heading off in all directions across the barren plane and some (internal part) passing through the (two overlapping and synchronized) trains – just imagine a doughnut getting fatter and fatter until the hole in the middle disappears. It appears to be a massive train-crash but, as already mentioned, the ultrasonic waves just take it in their stride(?) and they are not permanently affected (this is interference again on a grand scale). During this part of its journey, the interior of the original train on the track changes rapidly, with the radially converging doughnut sweeping through the interior of the original train. Nothing prevents the doughnut from passing through itself and once the (two overlapping and synchronized) trains have travelled sufficiently far the doughnut (internal) surface passes out of the carriage walls of the original train on the track. Now, the front of the first doughnut to be created is almost level with the front of the first train. Where this happens along the track is known as the transition from the Near Field to the Far Field. In the Near Field the overlapping doughnut is causing havoc (well, interference) and the shape of the train is changing rapidly. Inside the Near Field is not a good place to judge when the train arrives. In the Far Field the train has a stable shape again and is more or less like it was before it left the station. However, there are still waves heading off across the barren plane rather like an expanding balloon centred on the station where the train track exits.

Did I mention that every part of the (two overlapping and synchronized) trains generates its own doughnut so there are a continuous stream of doughnuts being generated – it's rather like a sonic boom wave - only in reverse.

Odd fact #5: I know I said that one, actually two trains get generated in the station but if the medium conveying the ultrasound is a solid (the train tracks) then more than one/two trains can be generated. This is because solids support both compression waves and shear waves. By contrast gases and liquids only support compression waves so trains running on gas or liquid tracks are somewhat simpler to understand. It gets worse because shear waves are transverse waves, (molecule motion is perpendicular to the direction of travel of the wave - like light) so the carriages of the shear-train can lie perpendicular and vertical to the train track (uncomfortable for passengers) or lie perpendicular and horizontal to the train track (much more comfortable) or, thanks to interference, a combination of vertical and horizontal (a bit like the pendulum tilting trains). Imagine when the one/two compression trains are created it is also possible to have one/two shear trains also created. Even worse is that shear-trains always travel slower than compression-trains (about 66% of the compression-train speed) and all the trains can interfere with each other and all the trains can diffract.

It's a complicated mess.

Odd fact #6: It is possible for compression trains to be converted into shear-trains and vice-versa. Mode-conversion generally happens at a free surface or an interface between two materials and there is an angular dependence upon the conversion.

Has the hallucinogenic drug helped you understand this weird train(s)?

With me so far?

Let's just have a reality check and make contact with physics.

- The creation of two waves in the station is a result of continuity of the medium through which the train is travelling (continuity of the tracks for the trains); or d'Alembert's solution to the wave equation in mathematics.
- In solids you can get compression-trains and shear trains generated. They can convert between each other.
- What happens at the buffers is reflection.
- When the trains overlap and add and subtract – that is interference.
- Having two or more overlapping and synchronized trains leaving a station – that is a common feature of ultrasonic transmitters called reverberation. Reverberation is generally responsible for ringing in the output of a transmitter and it is also responsible for creating a resonance in the transducer.
- The doughnut splitting off the train is diffraction.
- The business about the overlapping doughnut waves and where overlapping ends - this is where the Near Field (Fresnel diffraction) and the Far Field (Fraunhoffer diffraction) transition is found.

- The position of the Near Field depends upon ratio of the carriage length (wavelength) to the width of the train.

Engineers are always coming up with novel ways to get transmitters (stations) to emit the kind of ultrasonic waves (trains) they want. They must keep in mind what goes on in the train stations – if they do and they remember to use trains that have travelled more than the Near Field distance then everything can be hunky-dory.

Simple. All problems over – if only it was so simple.

In many applications of ultrasound, particularly sensing/measurement applications (instead of ultrasound treatment applications generally involving power ultrasound) the train has to go into another station so that its arrival time can be measured.

If only life were so simple that the train passes through the station unaltered – well it is possible if a non-mechanical method is used for sensing the ultrasound, such as optical or electro-magnetic sensing, but even then the mechanical things associated with the optical parts (for example: light source, lenses, photo-sensors) can cause problems if they interact physically with the ultrasound. Naturally, optical detection is impossible unless there is a transparent material through which the waves can travel (glass, some polymers, water or air). Cambridge Ultrasonics has been a pioneer for many years of the use of optical methods to render visible ultrasonic waves and our web-site shows photographs and video of ultrasonic waves travelling. We use our state-of-the-art equipment to help solve problems in clients' ultrasonic systems.

Let's assume the train arrives at another station – a mechanical ultrasonic receiver. Examples might be: a receiver having a piezoelectric element to convert mechanical displacement into an electrical signal; or in air, a microphone where a thin, lightweight membrane moves that is stretched over a rigid metal cavity, thereby changing the electrical capacitance of the cavity. What happens?

Actually, what happens is the reverse of what happened in the transmitter-station.

Let's get the trains going again. As the two overlapping and synchronized trains (let's forget shear-trains, diffracted trains and assume the trains are in the Far Field for a moment) reach the outside of the receiver-station they create another doughnut on the skin of the train. Part of the doughnut travels in the opposite direction out of the station and this part can hopefully be ignored but part of the doughnut travels into the station causing havoc because stations are usually shorter than the Near Field distance. This station is also equipped with buffers and the train and doughnut rebounds from them. In fact the station may have something like doors into the station so that some of the train rebounds off the doors as the train arrives but, if part of the train gets through the doors, the train can rebound off the inside of the doors and then travel back to the buffers and then back to the doors – creating a reverberating train. Some stations don't have buffers but have a long length of track laid in sand so that the train slows down and stops – actually the train does not slow to a stop it just gets smaller and smaller in height until it disappears. However, it's popular to make receiver stations as short as possible so that reverberations and internal diffractions are less noticeable and therefore less of a problem.

What does the receiver-station detect and when does the train arrive?

Unfortunately, the receiver station converts all the action of this train-crash into an electrical signal. To cope with this messy signal we need to choose a criterion for judging when the wave pulse (train) arrives and we are at liberty to choose whatever we like. Part 3 of this series of articles described a criterion that is sometimes used in flow-meters for oil and gas (ETC) but another commonly used criterion is: when does the earliest part of the wave arrive that is greater than a threshold value– I'll refer to it as the first-time-of-arrival (FTOA).

Let's consider two criteria for judging when this train arrives: FTOA and the end of the third carriage (ETC).

FTOA: What arrives in the receiver-station is a mess that is more like a train-crash and train-crash debris keeps bouncing around inside the receiver-station sometimes taking a long time to be absorbed or to escape. Consequently, FTOA has the big benefit that it ignores nearly all of the train crash and just detects the FTOA. Provided you are happy to ignore all information arriving after the FTOA until the train crash has died away then FTOA is simple and you can use transmitters and receivers that make chaotic, weird trains. These kinds of transmitters and receivers are generally cheap with high sensitivity, which are two big advantages for inclusion in a product. FTOA is used in applications to measure thickness, in concrete for example, there is a well-known test that measures the speed of sound (compression speed) in concrete and relates elastic wave speed to the strength of the concrete. I suspect FTOA is also used in the ultrasonic parking sensors found on most modern road vehicles.

Problems with FTOA: there are problems if any information of interest arrives after the first rising-edge of the pulse. For example: say the train passes a hole in the track along the way and the idea is to find the location of the hole. Some of the train will be unaffected by the hole and this part will get to the station first and start the train-crash going but some of the train will be affected by the hole and it will be partially reflected and diffracted by the hole – in solids it will also be mode-converted into a shear-train if the initial train is a compression-train. Shear-trains always travel slower than compression-trains so a shear-train will always arrive at the receiver-station after the compression-train. Provided the receiver-station can detect both compression-trains and shear-trains then there will be two pulses to detect with different arrival times for each pulse. In fact, the difference in time can be exploited to calculate where the hole is along the track. This method of flaw-detection is known as time-of-flight diffraction (TOFD) in the non-destructive examination business (NDT or NDE).

Unfortunately, the receiver station is probably a mess after the first compression-train arrives and TOFD has to suppress information during this time so that its FTOA criterion does not try to detect the second shear-train in the middle of the compression-train crash. This puts a restriction on the range of TOFD – the hole cannot be too close to the receiver.

An alternative is to use a separate shear sensitive receiver-station and hope it is totally insensitive to compression-trains.

What about ETC? Either a long train (at least 3 carriages) is expected in ETC or a train-crash in the transmitter or receiver or both is anticipated. If a long train is transmitted through the oil and gas then either the transmitter-station allows reverberations or a long train is deliberately synthesized. Generally, transmitter-stations and receiver-stations are identical in ultrasonic flow-meters because the ultrasound pulse is transmitted to travel in both directions (at different times) so the transmitter becomes the receiver and vice-versa when the direction of ultrasound travel is reversed. Consequently, if the transmitter-station creates a train-crash then the receiver will compound it with an even longer crash of the received train. ETC transducers sit inside apertures or ports in the otherwise smooth bore of the pipeline. Generally, the construction is of heavy-duty stainless steel and the transducers are recessed in the ports. All the metal corners around the port are perfect for causing diffraction and any irregularity in the flow of the oil or gas that causes the train tracks to shift direction a little (or a lot!) and hit these port/aperture corners is very likely to result in diffracted waves being generated from the port openings and diffracted waves interfering with the train arriving at the receiver-station. Interference from diffracted waves causes the main train to change its shape and adversely affect the measurement. This is serious for ETC because the method is fragile in its choice of criterion "when does the wave arrive" (see Part 3 for details).

What can be done about this chaotic station business? Can we design a station to generate or receive trains without all the chaos? The short answer is that chaos cannot be eliminated but we can mitigate the effects.

- It is possible to introduce mechanical damping to reduce reverberations. Damping also reduces the mechanical efficiency of the transducer, reducing the pressure in the ultrasonic wave generated by transmitters, which can adversely affect range of sensing.
- Make the length of the transducer as small as possible – to make the time delay in reverberations short – this is easier to do in receivers than transmitters.
- Aperture diffraction cannot be eliminated, although it can be reduced by putting absorbing materials at the aperture. Diffraction from ports can be reduced by careful mechanical design, basically keeping apertures away from waves.
- It is also possible to make receivers that are preferentially sensitive to compression-waves or shear-waves for use with solids.
- Finally, signal processing can be used to identify when a specific shape of signal is detected at a receiver. This will be considered in a later Part of this series.

The important lesson to learn is that the design and construction of ultrasonic transducers is very important indeed to making the judgement of when does an ultrasonic wave arrive. Making this judgement requires careful consideration of the many factors that can influence the shape of an ultrasonic pulse.

## Part 5 - Non-Linear Propagation.

Most of the theory of ultrasonic waves assumes the waves have very small amplitude, in fact, infinitesimally small amplitude – that's the last non-zero value before zero amplitude. Infinitesimally small waves are much simpler to understand mathematically because the equations are linear - the only problem is that nobody uses infinitesimally small waves because they are too small! The big advantage of linear equations is that their solutions obey the superposition principle: two different solutions can be added together to make a third solution. Interference relies on linear superposition of waves. Practical applications of ultrasound use waves of finite amplitude and that means we have to be prepared for non-linear effects.

Those of you reading who are feeling punch-drunk will probably have read Parts 1 to 4 of this series already and you may be growing tired of my analogy between trains and ultrasonic waves – resulting in some weird train journeys. If you have not read parts 1 to 4 then just remember a wave is a train and I will sacrifice our perception of real trains to bend the analogy so that trains behave as waves. You have been warned! Expect weird and you will not be disappointed.

How does non-linear propagation affect when a wave arrives?

Before answering that question, let's get more familiar with non-linear behaviour in a material by an example – the material is air. Most people have used a pump to inflate a bicycle tyre; some of you, like me, will have put your finger over the open end of the bicycle pump to seal it then tried to push-in the handle of the pump. To begin with it's easy to push and the handle moves in a reasonable distance, compressing the trapped gas. Then it gets progressively more difficult and eventually it becomes virtually impossible to compress the gas any more.

You just experienced the non-linear compression of air and exactly the same thing happens with ultrasound.

Consider for the moment a small volume of air that is fixed in space; as a wave goes through it there is a period of compression followed by a period of decompression or rarefaction and, eventually, once the wave has passed through completely, the pressure returns to ambient. While the wave was passing that little volume of air experienced compression followed by decompression – just like in the bicycle pump. On the compression cycle it got disproportionately harder to compress the molecules of air as the pressure increased while on the decompression cycle it got disproportionately easier to decompress as the pressure decreased. When air is compressed the air molecules are closer together, they collide more frequently, consequently, strong inter-molecular forces are active more frequently. The increase in frequency of collisions causes the speed of sound to increase as the pressure increases.

Let's recap here: the ultrasonic wave travels faster during the compressive half-cycle and slower during the decompressive half-cycle. If we start transmitting a strong sine-wave in air and place a good quality microphone close to the transmitter then the microphone will register the sine-wave. However, as the microphone is taken farther away from the transmitter the sine-wave gradually turns into a saw-tooth wave and eventually a triangular-shaped wave with a vertical leading edge.

Let's de-contextualize this physics and talk analogy-trains.

OK, up to now all the trains were infinitesimally small in height – very difficult to sit inside, very cramped. Now let's talk about trains that you can sit inside. Let's assume you paid for a premium ticket with a fancy reclining seat and you try to enjoy a sleep on the train by setting the chair to maximum reclining position. The train leaves the station, the seat is reclining and comfortable – there is a collision inside and just outside the station, we'll gloss over that (see Part 4) - but after a while things settle down.

Only things don't settle down.

As you travel farther away from the station, you start to notice the comfortable reclining seat making itself less and less reclining, in fact before long you find yourself sitting bolt upright so that the chair is very uncomfortable. The chair seems to have a mind of its own and it has also grown in height and the top is poking out of the top of the carriage. All this is very strange and annoying – not a very comfortable chair considering you paid extra to have it. As the train starts to travel farther down the track you observe that the top of the chair is getting progressively smaller and the reclining angle is less vertical and more comfortable again until the chair is quite small in height. You look around and notice that all the chairs are adjusted to the same reclining angle and they are all small – even those that are not premium seats. You feel cheated: you paid for a premium seat but first of all it adjusted itself to become vertical and very uncomfortable then, as the train travelled farther, the seat became a standard, small, economy seat.

You just experienced a non-linear train.

You take another train journey and after the chaos of the train station you are in the region where strange things happened before to your seat. This time, the train crosses-over another track and, disastrously, there is a second train on the crossing at the same time.

The two trains collide!

Well, if the trains were infinitesimally small trains then it wouldn't matter too much: there would be interference (see Part 4) but the trains would pass through each other and your train would return to being what it was like before the crash. But on this occasion the trains are not infinitesimally small and now something weird happens: there is still interference but at the crossing you find one of the passengers of the second train is suddenly sitting in your in lap – what the hell is going-on? Only it's even more weird because the person that has suddenly appeared in your lap is, firstly, quite small, secondly, the person is now travelling in your train instead of their own and, finally, they seemed to be locked in a trance. The other out-of-body experience is that a small part of <u>you</u> is now travelling in the other train in an entirely different direction; the other small you is behaving like your new friend sitting in your lap. This is the process of mixing of two waves.

Was that another hallucinogenic experience or what?

Of course, it's not just one new, small person that has been mysteriously transplanted into your train, its every person (and part of the train) that overlapped in the collision and they are positioned in your train where they were when the overlap of the two trains happened – does that make sense?

It's time for one of those reality checks with physics about non-linear effects:

- The rules of interference are broken and simple addition is no longer sufficient to describe what happens although interference does still occur. The additional effect is non-linear mixing.
- Mixing of two trains/waves happens so that each wave develops a component that was formerly a part of the wave with which it collided (mathematically it is multiplication - contrast multiplication with addition which happens during interference). Mixing results in making new waves that are either the sum or the difference of the frequencies of the two mixing waves.
- Even a single sine-wave can self-mix if it is sufficiently strong; the result is that harmonics of the wave are generated (addition of the same frequency – subtraction results in zero frequency). When the reclining chair gets steeper and higher, it is because non-linear propagation results in the creation of harmonics of the transmitted wave and the odd harmonics (3 x frequency, 5 x frequency etc) combine to turn the sine-wave fundamental (1 x frequency) into a saw-tooth and eventually a sharp, steep triangular wave.
- If harmonics have already been generated then mixing occurs between every conceivable combination of pairs of waves – resulting in very many different frequencies and a veritable cacophony of frequencies spreading out to high harmonics.
- Ultrasound is absorbed as it travels through most materials and the rate of absorption nearly always increases with frequency ($frequency^2$ actually is closer to reality). When a wave contains many harmonics, the higher frequencies are absorbed more quickly than the lower frequencies so that's why the seats shrink in height until there is only the fundamental frequency propagating but at low amplitude (small enough that non-linear effects are not significant). Overall, non-linear propagation results in a much greater attenuation of a strong wave than a weak wave. Not surprisingly, the stronger wave never becomes weaker than the weaker wave (provided they travel along identical paths), rather the stronger wave becomes increasingly like the weaker wave.
- The degree to which a train track (material through which ultrasound is travelling) generates non-linear effects depends upon the track/material. Approximate mathematical descriptions of finite amplitude waves use the parameter b/a, where a is the parameter for linear waves and b is the first non-linear parameter (in a potentially infinite series of parameters with ever-decreasing values). The magnitude of the b/a parameter is a measure of how readily non-linear effects are generated in

a particular material. For example the b/a parameter for water is larger than for air, consequently, non-linear effects are more readily observed when ultrasound is travelling in water than in air.

- Non-linear effects happen in the medium through which the wave/train is travelling; the more the wave travels through the non-linear medium the more the non-linear effects occur. This makes non-linear effects progressive with distance.

I hear you asking, “So what? What have non-linear effects ever done for me?”

How about if I first answer a closely related question firstly, “What have non-linear effects ever done to me?” If you are male and aged less than about 40 years then read the first example below carefully.

Here are a few examples of non-linear effects and products using them:

1. This is an example of where non-linear effects can be a problem. Medical ultrasound scanners use phased-arrays (relying heavily on interference) to make ultrasound focus at a point. A phased array is a collection of ultrasonic transducers; the transducers behave as both transmitters and receivers (train stations – for departure and arrival) with each transducer driven independently. Receive-signals from each transducer are also processed uniquely but eventually all signals are added together to create an image of the region being probed by ultrasound. Generally, the same shape of electrical signal is used to drive every transducer but, by changing the timings of the transmitted signals, the focal point can be positioned anywhere at will. Since the timing parameters are controlled electronically by a computer, the position of the focal point can be controlled by the same computer. The focal point is where the most intense ultrasound is found, consequently, any reflecting object at that point will produce the strongest echo, much stronger than echoes from elsewhere in the medium. By collecting echoes and scanning the focal point in the medium, an image can be made of the inside of an optically opaque medium. If the medium is the human body then an image of the interior of the body can be made. The problem with phased-arrays is that non-linear effects can happen at the focal point because the ultrasound is most intense there. Harmonics are generated at the focal point causing the ultrasonic pulse to become sharper with a more intense pressure than anticipated (more intense than simple interference predicts by simply adding together the pressures of the ultrasonic pulses from the individual transducers). High pressures can cause local heating and/or cavitation; the latter can result in sonochemical reactions. All these effects can happen inside the human body and they are bad for it but the most problematic region of the body is the brain of foetuses. Apparently, the brains of baby boys are particularly sensitive to temperature rises and it is believed that a temperature rise of as little as $2^{o}$ C can cause changes in the brain and changes in the eventual personality and behaviour of the child/adult in later life. (“What has non-linear ultrasound ever done to me?”)
2. Ever been to a rock concert and tried to listen to the music? I hope I am not alone in often only hearing a high-pitched screeching sound as well as the thump in my stomach from the bass drum or bass guitar. What made the screeching sound? One possibility is that non-linear propagation in the air of the intense sound waves

generates harmonics of the original sound turning it into a cacophony of high frequencies.

3. Ever listened to an opera singer from a distance of a several metres? The sound can sound screeching too, like at the rock concert described above. Get close to the singer (less than 1 m) and the sound is louder but the screeching disappears – this is a characteristic of the progressive nature of non-linear effects.
4. There are some commercial loudspeakers that exploit non-linear propagation in air. The loudspeakers are of normal diameter (20 cm to 50 cm) and they are generally thin and flat. They launch ultrasound at two frequencies close together at around 60 kHz. Because the ultrasonic wavelength is about 5 mm at 60 kHz and the diameters of the loudspeakers are very much more than 5 mm, the beam of ultrasound from the loudspeaker expands very little as it travels through the air (like a search light) and this helps to confine the sound that is made by the loudspeaker – this search-light beam is one of the big advantages of the loudspeaker. There are two ultrasonic frequencies present so they mix to make (at least) the sum frequency and the difference frequency. The sum frequency is also ultrasonic (about 60 + 60 = 120 kHz) and therefore inaudible but the difference frequency could be audible (say 61 – 60 = 1 kHz) and this difference frequency is the component that is exploited – the difference frequency is arranged to be audible and it is arranged to be the speech or music or audio sound to be made audible by the loudspeaker. A special electronic unit is needed to create the two ultrasonic frequencies. The characteristic of this loudspeaker is the search light beam of audible sound it makes. This type of loudspeaker has been used in museums and at point of sale locations in shops and stores so that a person strolling around will walk into the ultrasonic beam and hear audible information related to the exhibit or product nearby. The audible sound is created by the air through which the ultrasonic beam travels.
5. There is a sonar system (parametric array) that makes use of non-linear propagation by sending out two relatively high frequency sonar beams in the water in the same direction. Due to their high frequencies the sonar beams are rapidly attenuated and the projected beam has a range of about 1 km that is too short to be used directly for sonar. However, like the beam in the loudspeaker in the previous example, the two collinear sonar beams are also narrow. Importantly, the two narrow beams interfere in the water to make a line of standing waves. The anti-nodes (places of large amplitude) behave as a line-array of sources in the water over a distance of about 1 km. Each anti-node generates a low difference frequency by non-linearity of the water. Low (difference frequency) waves from all the anti-node sources travel out in all directions but in one direction (perpendicular to the line of standing waves) the waves interfere to make a narrow-beam, low frequency sonar source (effectively a phased array), which is virtually impossible to create in any other way. Low-frequency sonar has a long range, which is desirable for sonar, but it also has a narrow search light-like beam because the array is 1 km long and has many sources along its line. Low frequency, narrow-beam sonar is very desirable for sonar because it has long range and the narrowness of the beam helps greatly in localizing reflective targets.

So far I have discussed non-linear effects in the material that conveys ultrasonic waves – the train tracks. It's not the only source of non-linearity. Here are other sources:

- The power amplifier or electronic circuit used to drive the transmitter.
- The transmitter itself.
- The receiver.
- The amplifier used to amplify the signals from the receiver.

I don't propose to dive into the electronic circuits in this discussion but let's consider non-linear effects in transducers.

Piezoelectric devices have tensor (matrix) equations to describe the piezoelectric effect but they are all taken to be linear. True piezoelectric materials have a chemical crystal lattice that is non-centro symmetric – this is the all-important quality for a crystal to be piezoelectric and pyroelectric. Most commercial materials are ferroelectric (like ferromagnetic materials) and can be polarized by an electric field and polarized ferroelectric materials behave like piezoelectric materials due to rotation of the crystal lattices. If the crystal lattice is sufficiently rotated then the piezoelectric behaviour stops being linear. Alternatively, if there is a small crack in the crystal and ultrasonic vibration causes the crack to open and close then the elastic properties of the crystal will become non-linear (this is true for any solid material that is carrying an ultrasonic wave)

Condenser microphones comprise a hollow, sealed metal or plastic case having a small aperture and a thin polymer membrane stretched across the inside of the aperture. Within the microphone body there is a small volume of air. This should remind you of the bicycle pump that was first introduced as an example of non-linear material behaviour. Most microphones have resonances, at which the amplitude of vibration can become high, these are favourable conditions for non-linear effects to occur.

Let's not forget the fundamental question of interest in this discussion. How does non-linear propagation affect judgement of when a wave arrives?

Non-linear propagation of ultrasound (or sound) is a real shape changer: sine-waves get turned into triangular waves and back into low amplitude sine-waves as they travel through the propagating medium (air, water, the human body, steel etc…). If the material is also dispersive (see Part 2) then the harmonics generated by non-linear effects travel at different speeds from the fundamental frequency and can be variously absorbed. The different components remaining and arriving at a receiver interfere and result in changes of shape of the ultrasonic waves that can be difficult to predict. Changes in the propagating medium (say in oil and gas flowing in a pipeline) result in time-varying changes in the shape of the ultrasonic waves arriving at the receiver and this affects the judgement of when the wave arrives, leading possibly to errors in measurement of the flow-speed.

Finite amplitude waves are always used in real-World applications; consequently, non-linear effects are always present to some degree. When there is a problem with signal-to-noise ratio in a receiver signal, the first approach to improve performance is generally to make the transmitter signal stronger – but this will also increase non-linear effects, changing the shape of the wave pulse, which may not be desirable.

The shape that a wave pulse has when it arrives at a receiver can be very important in deciding when the pulse has arrived. Non-linear propagation changes the shape of the waves which can affect the judgement of time of arrival. The progressive nature of non-linear effects means that changing the strength of a transmitted signal can unexpectedly change the shape of the wave with distance. For example, in air, at a frequency of about 50 kHz, a sine-wave is turned into a triangular-wave over a distance of about 1 m from the transmitter; after about 15 m the wave returns to a low amplitude sine-wave due to preferential absorption of higher frequency harmonics of 50 kHz.

Next time you are on a train and you find a small person sitting on your lap … well, you will know what has happened.

## Part 6 - Scattering

Ultrasonic scattering lies at the very heart of all applications of ultrasound that involve inspection or imaging. However, scattering is generally unwelcome in other applications, for example: when using power ultrasound, scattering disturbs the ultrasonic beam resulting in sub-optimal processing; in communications, scattering results in multi-path fading.

If you have already read Parts 1 to 5 of this series then I want to congratulate you on your determination! My analogy between trains and ultrasonic waves, resulting in some weird train journeys, could cause confusion (strangely, it is an attempt to illuminate the subject!). If you have not read parts 1 to 5 then the important thing to keep in mind is that a wave is a train and trains have to be distorted so that they behave as waves. You have been warned! Expect weird and you will not be disappointed.

Let’s talk trains.

A train is travelling along a track and it hits a rock or boulder on the track. What happens? In ultrasonic terminology the wave is scattered by the rock. Let’s convert what happens to the wave into what would happen to a train. What happens depends upon the relative size of the train and the rock – it sounds reasonable for a train too. Three types of scattering occur:

1. If the rock is large compared to the carriage length then the effect of collision/scattering is massive. The train is reflected – the train does not bounce off the rock and lurch backwards as a real train might – every part of the train runs forward to strike the rock and to be reflected. It’s just like the buffers in the station when the train starts (see Part 1).
    a. Transmission and reflection. Depending upon how hard the rock is in relation to how hard the train track is, some of the train will be reflected but some can be transmitted. This is weird because it means the train splits into two parts: one part passing through the rock and other part being reflected from the rock. So one train turns into two trains and although they remain the same number of carriages the height of the two trains can change depending

upon the hardness of the rock (actually its acoustic impedance which is closely related to hardness).

b. The part of the train that is reflected heads off in a direction that depends upon the surface of the rock: if the rock is perfectly flat and its surface normal is pointing along the train track in the direction opposite to the train's direction of motion then the reflected part of the train heads back up the track in the opposite direction but if the surface normal is pointing away from the track then the train travels off obliquely at twice the angle between the track and the surface normal to the rock. If the rock face is faceted then there is one reflected train for every facet – if could be a lot of trains! Where these trains overlap then they interfere (if the trains are all very small in size) if the reflected trains are all large then they will interfere with non-linear mixing. This quickly becomes messy!
c. The part of the train that is transmitted changes speed to a speed determined by the rock and that means if the incident train (the original train before striking the rock) strikes the rock obliquely on a surface then the train gets abruptly changed in direction by an amount that depends upon the ratio of the train's speed before striking the rock and the train's speed in the rock. In wave terminology this is called refraction. For a multi-faceted rock surface, refraction happens at every surface and the refracted waves all interfere in the rock by interference and, possibly, by non-linear mixing. This also quickly becomes messy!
d. If the rock is solid (and which rocks are not?) then there are compression trains and shear trains and in crystalline rocks the two orthogonal types of shear train can travel at different speeds, both of which are different from the compression train speed. Even messier!

2. Rock is much smaller than the train. Everything that happens when the rock is large still happens except the magnitudes of the effects are all much smaller and, in many instances, they are negligibly small (especially trains internal in the rock). There is one new effect: the vast majority of the train continues onwards unaffected. If there are lots of small scatterers then reflected/scattered trains can combine to make a larger effect but now interference can happen between all the scattered trains to make a larger effect.
3. Rock is about the same size as a carriage. Again, everything that happens when the rock is larger than the train happens but at a smaller magnitude and there is still a significant amount of the train that travels onwards unaffected. But now the scattered components are significant in magnitude (compared with the unscattered train) and these scattered trains are created at every rock that is encountered. If there are many rocks, suddenly, there are trains going in every direction and multiple scattering can happen too. It is complete chaos! This is called multiplicative scattering. The degree of chaos depends upon how different are the material properties of the rock and the tracks (between the rocks).

In fact, nearly all materials have a microscopic structure at small sizes. The human body along with animal bodies are made up of cells and groups of cells so that at high ultrasonic frequencies multiplicative scattering happens in medical ultrasound. Most metals crystallize to form crystallites or grains with boundaries; if the crystal is anisotropic then the crystallites will have random orientation of their planes at boundaries resulting in scattering. Cast iron is an example. Alloys, such as the various forms of steel, often have non-uniform distributions of their alloying components around their grains often with carbon located at grain boundaries. When the ultrasonic wavelength becomes as small as the grain size multiplicative scattering happens.

The effect of multiplicative scattering is the same as turning a single, well-defined train of finite duration (possibly short) and clearly defined carriages into a longer pulse where the carriage length has to some degree been randomized and with a profile that is controlled by how many rocks have been passed along the track. If only a few rocks have been passed and if the rocks are not too different from the tracks (in terms of material properties) then a train of some kind will be detected. However, if there are many rocks and if they are significantly different from the track then the emerging train is so small and so long that it may be undetectable.

If a plane wave is launched at a transmitter (a train where the track is much wider than the carriage length – admittedly another weird train!) then what escapes from all the rocks after multiplicative scattering is lots of trains where the tracks are about the same width as the carriages. The scattering material does not support plane waves.

Scattering is often described as one of the mechanisms of ultrasonic wave attenuation along with absorption. I am not happy with this linkage. Absorption converts the ultrasonic (mechanical) wave into heat by a process that is irreversible. Scattering disrupts the coherence of a transmitted ultrasonic wave: its spatial coherence and temporal coherence are disrupted. Does it convert the wave into another form of energy such as heat? The answer is no. Experiments show that if two matched plane wave transducers are used to measure the attenuation of a material that exhibits multiplicative scattering they will register a high degree of attenuation. However, if one of the plane wave transducers is kept as a transmitter but an array of small aperture transducers is used as the receiver, having the same total area s the first planar receiver and with each individual receiver having an aperture no greater than the wavelength of ultrasound used and if the signals from individual receivers are combined without phase cancelling, then a large signal is still detected and a signal much larger than was detected with the matched, planar receiver.

I believe that attenuation should be an intrinsic property of the material that should not depend upon the construction of the receiver.

This begs the question: what is attenuation?

## Part 7 - Energy Arrival

If you have not read Parts 1 to 6 of this series and you plan to jump straight into Part 7 then you are going to save a lot of time but bear in mind an underlying analogy between a pulse of ultrasonic waves and a train on a track.

I hope by now I have convinced you to give up trying to measure when a wave arrives! It is certainly a tricky question to answer and plagued by all the effects that can happen to derail your measurement.

More seriously, the practitioner wanting to measure when a wave arrives is advised to choose a criterion for judging when a wave arrives that is both practical to implement and takes account of the pitfalls described in earlier sections of this document. In general, I prefer to measure when maximum energy arrives but more of that later. Let's recap some of the problems of some of the popular methods.

First time of Arrival (FTOA)

1. The true first time of arrival measures the first detectable energy of the wave. This is a terrible time to make a measurement! Let's talk trains.
2. Say there is mist in the air or a sandstorm or a snowstorm or anything that introduces a random noise in the measurement. Almost every measurement of this kind is made electronically and there is always electrical noise present. Generally the largest component is from a resistor electrically close to the sensor – at the start of the amplification chain. This resistor makes random noise. You stand on the platform (where?) and look at a point on the track but there is mist or sand or snow swirling about. You think the train might be about to arrive.
3. Let's say you have a high speed camera watching the chosen point on the track set-up at that important point on the platform. Let's say you can go back to look at the footage of the train's arrival. When does it arrive?
    a. Is it when the snow starts to swirl because of the air motion ahead of the train?
    b. Is it when you can see the very first part of the buffers at the start of the train?
    c. Is it when the start of the driver's carriage arrives?
    d. What if the driver's carriage has an aerodynamic profile – does the train arrive when the driver is visible?
    e. All these events are shrouded in mist or sand or snow and it's difficult to tell when these events happen. It is hardest of all to judge when the very first part of the curved front surface of the buffers cross the point you have chosen on the track.
4. The problem with FTOA is that the noise in the measurement is large compared to the signal and that introduces a random value into the time that you measure for the FTOA.
5. The true FTOA, the very first detectable arrival time has a terrible signal-to-noise ratio and the noise creates a random error that can be unacceptably large.

6. Most so-called FTOA do not measure the true FTOA but measure when the driver arrives; some threshold voltage is set that is higher than say 5 standard deviations of noise. But this is not the first time of arrival!
7. Another problem is dispersion. FTOA measures the arrival of the fastest phase. If the purpose of the measurement is to assess the material then the fastest phase may be least sensitive to the material property of interest.
8. One way to improve FTOA measurement is to fit a curve of what you expect the wave profile to be – you know the shape of the train. Collect points on the wave where the signal-to-noise ratio is good, fit a suitable curve and extrapolate backwards to where the curve crosses zero. This should be the true FTOA.
9. FTOA is relatively robust against multiplicative scattering.

Last Time of Arrival (LTOA)

I include this more for my own sense of weirdness because I don't know of any applications of it!

What is LTOA – it is simply the trailing end of the wave – the last carriage to pass-by.

1. In principle, LTOA has all the same problems of FTOA except that you already know that the train has been passing-by.
2. Point 8 of FTOA can be used to find LTOA more accurately.
3. LTOA is strongly affected by both dispersion and multiplicative scattering.

End of the Third Carriage (ETC)

1. ETC should not suffer from the problems of random noise that FTOA or LTOA suffer. This is because the amplitude of the signal is strong and the event of the signal passing through zero is easy to detect and can be accurately determined.
2. ETC introduces huge errors if the measurement system makes an error in counting to three (wave crosses through zero or three carriages passing). This can easily happen at the beginning of the pulse for exactly the same reasons that FTOA has a problem.
3. ETC can also introduce errors when there is dispersion present or there is scattering. In oil and gas metering the ultrasonic beam is displaced by the motion of the fluid and this can result in the beam striking the port in the pipeline where ultrasonic transducers are located resulting in diffracted waves from the ports changing the wave profile.
4. Pressure pulses in the fluid can also cause problems and many ETC systems ignore pressure pulses by forcing results to remain static if there is a sudden variation in value.
5. ETC, like most arrival time measurement systems, has an arbitrariness in it that is both pragmatic and breath-taking in its boldness. It relies upon calibration, using large quantities of the fluid of interest passing through the measurement region, to convert that arbitrariness into reliable values.
6. Ultimately, ETC does not cope well with pressure pulses.

Arrival of Peak Energy (AOPE)

AOPE has been alluded to in other sections of this document but it hasn't been described in detail.

Let's talk trains.

Most trains arriving at a station start off small, then get bigger and bigger, reach a maximum, then get smaller and smaller until they disappear. Peak energy is simply the biggest point.

Waves are awkwardly bi-polar, meaning they have positive values and negative values: if these are ultrasonic waves then the positive corresponds to high pressure and negative is rarefaction (low pressure). After conversion to an electrical signal we get positive and negative voltages. This bipolar nature means that peak energy could correspond to the most negative point or the most positive point. It is important to take the magnitude of the signal to turn the bi-polar signal into a uni-polar - it is much easier to measure the peak point of a uni-polar signal.

A popular approach is to rectify and smooth the signal – it's popular because it is easy to do electronically but it discards half the signal so it introduces errors. I prefer to take the magnitude of the analytic signal or the Hilbert transform of the signal because it doesn't discard any information – unfortunately it is computationally expensive.

A more sophisticated approach is to launch a known wave that exploits the bandwidth of the transducers. I am fond of chirps – so were the guys that invented radar because they use it extensively. A typical chirp might start at a low frequency, say 20 kHz, and sweep in time up to a higher frequency, say 50 kHz, all in a sweep-time of 500 µs. The sweep-time is long enough to allow many cycles to be included. The frequency-sweep can be linear or logarithmic or many other mathematical functions. What would an analogous train look like? Think of a train where the carriage length gets progressively shorter along the length of the train. Alternatively, carriage length could get progressively longer. The important point is that there must be a known pattern in the carriage length.

This approach exploits that there is a pattern in the transmitted ultrasonic chirp. This pattern should still be present in the received ultrasonic waves and therefore in the electronic signal created by the receiver. Computers are good at finding patterns – algorithms such as match filtering or cross-correlation can be used. What happens when one of these algorithms detects a chirp is interesting: in principle, it should produce an infinitesimally sharp spike (remember the knife-width train passing through a station that was knife-width long resulted in a sharp spike for when the train arrived?), which would make it perfect timing measurement. Is this method too good to be true? Let's call this method MF-AOPE

1. AOPE is a robust method in finding time of arrival because the signal is largest and therefore the signal-to-noise ratio is at its best.

2. There is always only one point in the wave that is the largest so there is no ambiguity.
3. Some care needs to be applied in finding the largest point in the wave-pulse because it is the magnitude of the signal that must be considered – the peak energy could arrive when the signal is negative. The Hilbert transformation does a good job of helping to find peak energy point.
4. AOPE is sensitive to: dispersion, multiplicative scattering, interference, diffraction and transducer design. This may be a good thing if you want to measure material properties, for example, but many engineers regard it as a disadvantage.
5. MF-AOPE has a peak position that is determined by the autocorrelation of the chirp; the position of the peak is at the start of the chirp so it is a form of FTOA. However, all of the chirp contributes to where the peak is located and the signal-to-noise ratio is at its highest.
6. The height of the MF-AOPE peak is a measure of the energy of the pulse. It means that the more energy in the pulse the higher the peak. More energy can be supplied by either increasing the drive power or the duration of the pulse. It also means that the peak height can be used to measure the effect of attenuation along the transmission path.
7. MF-AOPE provides useful compression of the duration of the chirp: compression is the duration of the auto-correlation pulse main peak compared to the duration of the chirp. Compression factors of more than 5x are possible. However, this means that the method does not provide a perfectly sharp spike to measure when the wave arrives.
8. MF-AOPE produces an auto-correlation width that is equal to half of the reciprocal of the bandwidth of the original chirp. For the example chirp given earlier (20 kHz to 50 kHz) the bandwidth is 30 kHz so the auto-correlation peak width is 16.5 µs – this is not an infinitely sharp spike.
9. The main problem with MF-AOPE is related to the transducers. The method requires transducers to have high-fidelity for creation/reception of the chirp. For the example chirp given earlier the transducers must have a flat frequency response between 20 kHz and 50 kHz. This is difficult if not impossible to achieve. The wider the bandwidth of the chirp the narrower the peak of the auto-correlation of the chirp and the sharper is the detection of when the wave arrives.

The last point about MF-AOPE is of considerable importance because it connects with one of the laws of physics – Heisenberg's uncertainty principle. It's not surprising that it should because it they are both statements of the same thing and both go a long way to answer the question.

When does a wave arrive?

The answer is that we can never know more precisely than about the reciprocal of the bandwidth of the received signal. In principle it doesn't matter how we measure when the wave arrives – MF-AOPE is my personal favourite but the other methods have their advantages and disadvantages. None of them can hope to make the measurement more accurately than the reciprocal of the bandwidth of the received ultrasonic wave.

Let’s remember some early trains.

Remember the infinitely long train with carriages all the same length? You get to the station and the train is rumbling through it; you leave the station and the train is still rumbling through. You have no idea when the train arrived. This is a wave of infinite duration and zero bandwidth. The reciprocal of the bandwidth is infinity so the pulse that tells you when the train arrives is a flat line - there is no pulse. You cannot tell when the pulse or the train arrives - which fits with our observations.

Remember the train that was a knife-width wide? Take the Fourier transform of the equivalent wave and you get an infinite range of frequencies – the bandwidth is infinite. The reciprocal of infinity is zero, so the sharp spike is infinitesimally thin. We know perfectly when the train arrives.

I sense we are getting close to some useful advice about judging when a wave arrives. In ultrasonic applications, the practitioner is at liberty to choose a method that is convenient although I would recommend first trying MF-AOPE. However, remember that the accuracy of the measurement can never be better than approximately +/- ½ $(\text{bandwidth})^{-1}$.

For those in the oil and gas industry who scorn the above statement and claim a much better accuracy then remember that the accuracy statement above applies to a single, independent measurement. Their method often uses averaging over many measurements and it discards measurements that lie outside of an acceptable deviation from the running average. Effectively they are applying Baysian conditional probability to the result – that’s cheating!

Finally, remember in Part 6 on scattering there was a useful lesson to learn about what a receiver should be like. I recommended the use of an array of small aperture receivers and adding together the magnitudes of the signals. Provided each aperture is smaller than a wavelength then this receiver should be robust against multiplicative scatter and many other problems. An ultrasonic test that uses a plane wave transmitter seldom provides plane waves at the receiver. Receiver arrays should be the first choice in most applications.

Making ultrasonic measurements of arrival time has many concealed pitfalls. It is good to be aware of them.

Beware any train journey that is the same as the propagation of an ultrasonic wave!