Thermal and Residual Stress

A closer Look at Stress and Strain

Science

First we have to look a bit more closely into what the terms *stress* and *strain* actually mean. Yes, I did give you a simple and good definition way back in chapter 3 (and in a related science module) but we still need to look a bit more closely. Here we will do a few thought experiments to find out what, exactly, constitutes stress and strain - and what just looks like stress or strain but is something else. The relation between *displacement* and strain gives an example; it can be found in this link

I'm going to use the expressions "stress" and "strain" as more or less synonymous because if you know one you know the other. As longs as *only elastic deformation* occurs, stress and strain are simply related by elastic modules like Young's modulus, shear modulus and so on, or the elastic coefficients if you want to be fussy.

Since the topic is "*thermal* stress", I'm going to play around with temperature a bit. Let's heat up some steel, for example your sword blade.

Increasing the temperature of a solid will (almost) always increase its size - we have what is called **thermal expansion**. In a science module way back I have actually shown *why* there is always some thermal expansion and *how* one can actually calculate it.

The **thermal expansion coefficient** α puts a number on this material property. It is given by

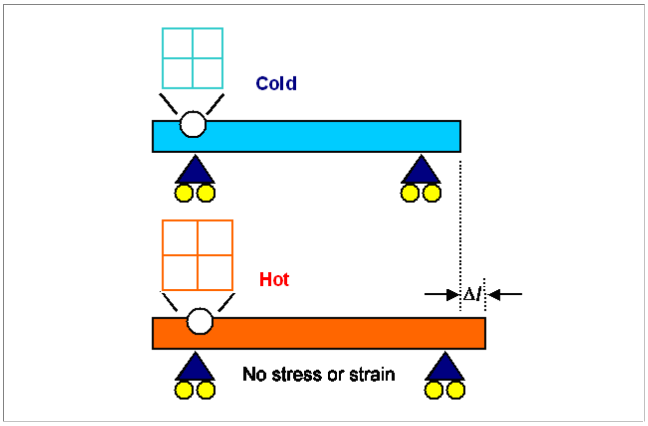
$$\alpha = \underbrace{\frac{l(T) - l(T_0)}{l(T_0)}}_{\text{Relative length change}} \cdot \underbrace{\frac{1}{T - T_0}}_{\text{Temperature difference}}$$

- So the numerical value of the thermal expansion coefficient is simply the *relative length change* caused by a temperature change of 1 K. Its unit is **[α] = K^{-1}**. It is a property of the bonds between the atoms and that means, as in the case of Young's modulus, that all (low-alloy) steels have pretty much the *same thermal expansion coefficient* as pure iron:

$$\alpha_{\text{iron}} = 11.8 \cdot 10^{-6}\ \text{K}^{-1} \quad \text{or}$$
$$\alpha_{\text{iron}} = 11.8\ \mu\text{m/m}\cdot\text{K}$$

The last number means that per **m** length an expansion of **11.8 µm** takes place if you increase the temperature by **1 K** or, same thing, by **1 ^{o}C**.

Now let's do our simple thought experiment. Take a rod of something and heat it up. It will get somewhat longer (and a little bit thinner) - provided you don't keep it between two ungiving restraints. Schematically it looks like this:



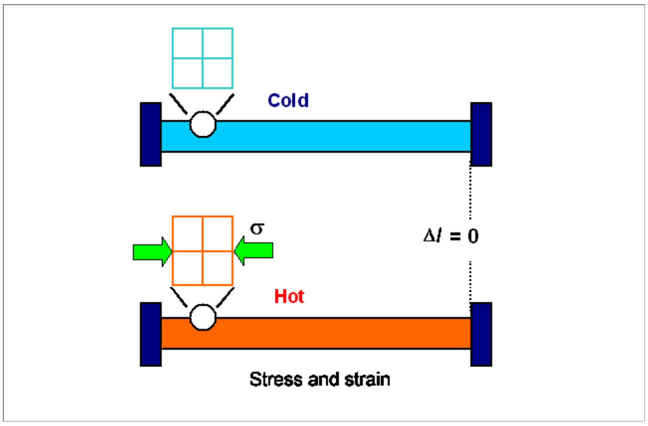
Thermal expansion without restraints.

Seen from the *outside*, you will have a *relative length change* $\Delta l_{\mathbf{therm}} = \alpha \cdot \Delta T \cdot l_0$.
When relative length changes were caused by stresses, i.e. force acting on the material, we have called that a *strain* ϵ, defined by $\epsilon = (l - l_0)/l_0$. So do we have something like a thermal strain $\epsilon_{\mathbf{therm}} = \alpha \cdot \Delta T$ here? Note that there is no stress whatsoever in our experiment. No forces were applied.
The answer is: *No!!!*. There is no strain or stress. You must be a bit careful about your frame of reference when invoking stress and strain. *You* look at your hot sword blade from the *outside* and compare it to some cold state. That's fine but there are other ways to look at this. Seen from the *inside*, or from the viewpoint of the *crystal*, there is no strain and no stress because it has the length it likes to have at the given temperature. In formal terms we have

- *Outside view:* $l_0 = l(T_0)$ = constant for all temperatures.
- *Inside view:* $l_0 = l(T)$ = increasing with T.

It should go without saying: As far as processes like deformation or fracture are concerned, the *inside* view is the one that counts.

Now we repeat the experiment but do not allow our rod to change its length. We just keep it between massive unmovable restraints. There can be thus no length change and there is thus no strain as perceived from the *outside*. In contrast, we now have plenty of stress. The rod is pressing hard on the unyielding restraints and since they are unyielding they press back with equal force or stress. It can be visualized like this:



Thermal stress and strain with restraints.

It is easy to calculate the stress in the sample. You have $\sigma_{\mathbf{therm}} = Y \cdot \alpha \cdot \Delta T$ because this is the stress you need to counter the thermal expansion, or, in plain words, to compress the expanded rod back to its original length.
In the *inside* view you have the same stress but also the *strain* that goes with it.
So, as a *first* conclusion, different temperatures between your specimen and the environment may or may not cause stress and strain in the specimen - it all depends on the "boundary" conditions.

From now on we will look at stress and strain exclusively from the inside view. This is very easy. All you need to do is to delve inside a crystal (use your brain microscope for this). Look at an elementary cell of the crystal. If is is exactly what it is supposed to be at the local temperature encountered there, you have neither local stress nor strain. If the elementary cell is not what it is supposed to be, you have stress and strain.
For iron, the elementary cell is always a cube. The dimension changes smoothly with temperature and abruptly when you have a fcc ⇔ bcc transition - but it is always well defined.

If you find that the elementary cell (EC) is not what it is supposed to be, you can draw a few conclusions from its shape (look at the old picture to "see" it):

- The EC is still a cube but either too large or too small. You have uniform compression or expansion (tension in all directons) , respectively.
- The EC is a cuboid (longer in one direction). You have uniaxial tension.
- At least one of the formerly square sides of the EC looks like a parallelogram now. You have shear stresses (and also normal stresses if the volume has changed, too).

Temperature Gradients Cause Stress and Strain

Now let's do another simple thought experiment. Take a cube of something. Take out a *smaller* cube from the interior (easy with your brain knife), heat it to some higher temperature, and put it back again. We are looking at a hot potato or whatever is hot inside and cold outside.

  - The hot core won't fit into its original place because it is now larger due to thermal expansion. You must either squeeze it to its old size, putting it under *compressive* stress and strain, or you must pull at the cold piece, elongating it in all directions by the proper amount to make the inside space large enough. In other words: you must put the outside under tensile stress and strain. Read up about stress and strain here, if you need a refresher.
    In reality, of course, for a hot core inside a cold shell, the stresses are shared between core and surroundings. The hot inside will be under compressive stress and the cold outside under tensile stress. If the inside is cold and the outside is hot (this happens when you put your sample into a hot furnace), it's the other way around. Note that the EC's are not what they are supposed to be. The ones in the hot core are too small for the temperatures there, the ones at the outside are too big, and in between it's messy anyway.
    The long and short of all of that is:

Thermal *gradients* in a specimen *always* cause mechanical stress and strain

  - "Thermal gradient" is a fancy word for "differences in temperature on a small scale", or $dT/dx \neq 0$ in math lingo for one dimension. In full generality for three dimensions the temperature gradient is a vector with the components dT/dx, dT/dy, dT/dz, or ∇T in shorthand.

- All of the above is tied to thermal expansion, the change of the lattice constant with temperature. However, we also might have expansion or compression whenever a **phase change** occurs. This is easy to see if we do another simple thought experiment
  - Take a cube of pure iron and heat it to 912 °C (1.674 °F). It will then be in the fcc phase. Take out a *smaller* cube from the interior (easy with your brain knife), cool it to 908 °C (1.666 °F), and put it back again. The dimensional change due to thermal contraction is so small that you can neglect it. Your slightly colder piece, however, will not fit because below 910 °C it changed to less dense bcc iron and thus expanded. Again you need to squeeze it, and so on, so it can be put back inside. What we conclude is:

Phase changes in only parts of the specimen *always* cause mechanical stress and strain

  - The specific volume or the density of different phases is generally different, and that means that *local* phase changes can also cause a lot of stress and strain. We use that, after all, to measure phase and TTT diagrams. The "local" is important. If *all* of the sample has changed to a different phase, there is no stress and strain anymore (assuming that there are no ungiving restraints).
  - For steel that is certainly true. Ferrite (or martensite) has a bcc type lattice and the crystal is thus by definition not as densely packed as the fcc austenite. Ferrite or martensite formation from austenite thus always *expands* the lattice. Martensite is also less dense then ferrite since it contains all this carbon that must expand the lattice somewhat on average. Of course, if we also have cementite in the mix, things get more complicated.

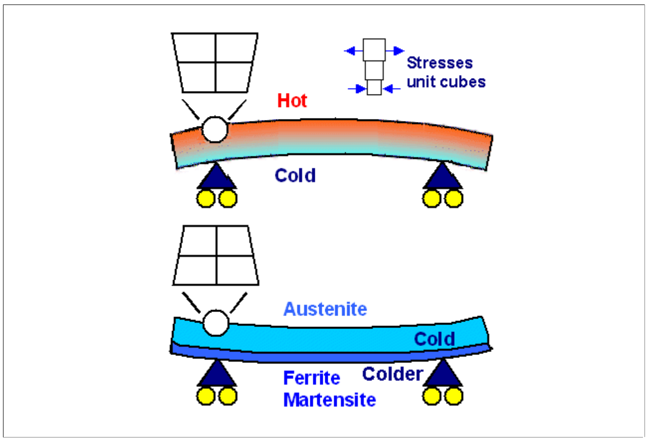
- We now have two mechanisms that can produce stress and strain and it is clear that both are related. If there are temperature gradients, the more quickly cooling parts of the sample will experience phase transitions sooner than the more slowly cooling parts, producing local phase changes and the strain / stress going with that.
  - This can lead to tricky situations. Assume, for example, that the "lower" part of your sample cools down more quickly than the upper part (you might imagine the edge of a katana sword blade during quenching). Let's list what could happen (shown schematically in the figure below):

    1. Let's assume the blade is *pure iron* and ferrite everywhere when you start cooling. It then will stay pure ferrite at all times. During cooling the upper (thicker) part is always hotter than the lower (thinner) part with the edge. The stress produced by the temperature gradient then always bends the blade *downwards* until the temperature is the same everywhere. The larger the gradient, the larger the bending. This is illustrated in the upper part of the figure below.
    2. Now let's assume that the blade is still *pure iron* but austenite everywhere when you start cooling. The austenite - ferrite transformation temperature then is first reached in the lower part. Ferrite formation thus starts in the lower part and moves upward. The "local phase change" stress is produced at the phase boundary where the lattice constants don't match.

The phase transformation austenite / ferrite by itself, while it takes place, leads to an *upwards* bending until the transformation is complete everywhere. This is illustrated in the lower part of the figure below. Note that you don't need large temperature gradients for producing major effects. It's enough if the lower part is slightly below the transformation temperature and the lower part slightly above. The "local phase change" induced bending is proportional to the difference in lattice constants between the two phases.
On top of that you still have the downward bending caused be the temperature gradient and thermal expansion. It is somewhat different from case 1 because the thermal expansion of ferrite is different to that of austenite - but you still have temperature gradient induced *downwards* bending.
The total bending effect will be given by the superposition of the two mechanisms. It could go either way, depending on "details".

3. Now let's assume that the blade is *homogeneous steel* with some carbon concentration, and that you start cooling in the austenite state. We have all the effects from case 2 with the added complication that the phase transformation is to some pearlite plus ferrite or cementite. The "local phase change" induced *upward* bending part caused be the phase transformation sweeping through the blade will thus have a more complex behavior. The *local* bending that we need to consider now, is proportional to the difference of the (average) lattice constants of the phases encountered *locally*.
4. Take case three and add *martensite formation* in the edge area where the cooling rate is fastest. This leads to major stress with respect to the still present austenite there and still some stress to the eventually forming ferrite / pearlite / bainite. In any case the martensite part would tend to bend the blade *upwards*.
5. Take case four but now for a *real* katana, with *soft steel* inside and *hard steel* outside, that is pitched into cold water. Extreme temperature gradients are produced. What will happen during cooling, considering that all kinds of phase transformations are encountered locally at different times / temperatures, is not all that easy to figure out.

- Once more: The total strain / stress produced at some given time in any part of the blade is obtained by summing up all those effects. Which way the blade will bend at any moment in time, what else might happen, and what the final shape will be, is far from being obvious.



Thermal stress and strain due to *only* temperature gradients or only local phase formation.

- In other words: it is easy to quench a hot blade in cold water or some other cold liquid, but not so easy to figure out beforehand exactly what is going to happen. And we are far from being done yet.
  - To make things a bit more complicated, it is necessary to point out that the trapezoids showing the deformation above may reflect the shape of real elementary cells of the crystal but must not be confused with a "unit cube" that has been deformed by the acting stress tensor. You *cannot* deform a cube into a trapezoid by one set of forces acting on its surface. You need to change forces over a sequence of stacked cubes or, in other words, the components of the stress tensor must vary in a certain way going from the bottom to the top as shown in the upper right hand corner of the figure.
    But that doesn't need to bother us here. We are not going to calculate anything here. It will be sufficient to deduce what is going to happen in a qualitative way.

Global, Local, and Microscopic Stress / Strain

- First we need to realize that all of the above is rather general. It would be true for perfect single crystals and for fully amorphous materials like glass. The silicon people, for example, dealing with rather perfect single crystals, can tell you a lot about stress and strain produced in silicon wafers by thermal gradients or by having a different phase (like silicon dioxide (SiO_2)) on one side of the specimen.
  On the other hand, a silicon wafer at constant temperature and not molested by other phases is *completely* stress and strain free, no matter what the temperature. Another way of saying that is: wherever you look inside the crystal you find a perfect unit cell of the crystal structure with just the right dimensions for the given temperature. Yet another way of stating this is to say that a perfect single crystal at constant temperature contains no *residual* stress and strain.
  - Now look at your sword blade or car. Dealing with iron and steel (or just about any metal), we *never ever* come even close to a perfect crystal. We have grain boundaries, dislocations, precipitates and God knows what other defects in there, typically at quite high densities. The density type and distribution of those defects will typically be different after the thermal treatment to what it was before.
    Now any dislocation has a stress and strain field surrounding its core. Precipitates of any size, grain boundaries, or just about any crystal lattice defects also have stress and strain fields in their immediate environments, no matter if the specimen is hot or cold or experiences thermal gradients. "*Immediate environment*" means that the stress from some defect is only noticeable close to the defect on a scale of some ten nanometers at best.

Defects produce stress and strain on a *microscopic* scale

- In the figures above I have shown what I called *local* stress and strain. It changes when you move across a specimen but on a macroscopic scale. The whole edge of your blade might experience local strain = strain in the edge = strain on a macroscopic or "*global*" scale. It is clear (at least in principle) how thermal gradients and what I now will call "**phase distribution gradients**" produce large scale if local stress / strain, and that the effects disappears in a perfect crystal as soon as the gradients are zero.
  - In the final product we typically only worry about large scale stress and strain. As long as the defects in there are uniformly distributed, their microscopic stress and strain fields average to about zero on large scales, on small scales it is a kind of background "noise".
    In other words: If your object has an easily recognizable different shape after the heat treatment (e.g. your blade is now bend), is has experienced large-scale stress and reacted by changing its shape. If its shape is not changed noticeably on a large scale, a formerly mirror-like surface might now be a bit dull, because the microscopic strain around defects produced during cooling makes it rough on a small scale.
- Generalizing a bit, we note: Defects in a crystal are the *only* sources of internal stress and strain. A perfect crystal not under any external constraints and thus not experiencing stress on its outside, is stress-free by definition. Thinking that through you realize that the *global* stress / strain state of a non-constrained specimen at a *constant* temperature can only be the sum of all the microscopic stress and strain around its defects. So sum up a lot of tensors and see what you get!
  - Haha. Nothing like a good joke among friends every now and then. Let's only look at two easy points:

    1. If you just look at the stress picture of a dislocation, you see just as much red as green, or just as much compressive as tensile stress. It's then easy to conceive, I hope, that the stress and strain fields of a random distribution of random dislocations essentially average to zero. For every compressed part somewhere, there is a stretched part somewhere else.
       That's also true in the spirit for all other kinds of defects.
    2. Generalizing from that, we can deduce that large-scale or global *residual* stress and strain at constant temperature can only results from an *non-uniform* distribution of defects (or phases).
  - Time to look a bit more closely on what really happens when cooling down a sword blade.

Residual Stress and Strain on Large Scales

- Sooner or later thermal gradients disappear. Your specimen or sword now has the same temperature and the same phase(s) everywhere. Then all large-scale stress and strain caused by temperature gradients and local phase differences during its making should be gone. Ignoring the microscopic stuff coming from defects, your specimen should have its old shape and there should be no **residual stress or strain** left in your blade. Is that true?
  If that would be always true, I wouldn't need to write a lot more in this module. Major complications come from the fact, that the quiet little assumption I made all the way up there, is quite often not correct. It is often *not* true that *only elastic deformation* occurs as a reaction to stress induced by temperature gradients or local phase changes!
  Taking this into account, we might now distinguish between two basic cases:

- **1.** *Local stresses* produced during cooling by any one, or any combination, of the possibilities enumerated above will always (= for at all local temperatures and at all times) be *below some critical stress*. Then you will only find elastic strain that disappears if the stress causing factor disappears. To what extent will the stress causing factors disappear? Let's see by looking at the stress causing factors individually (i.e. assuming that they are the only ones in some specimen):

    1. *Temperature gradients* only. Eventually they disappear completely - and so does the elastic stress / strain caused by them. If you start with a straight iron blade (or, more to the point, a perfectly flat silicon wafer), you may observe substantial bending during cooling but always end up with a straight iron blade (perfectly flat silicon wafer) identical in shape to the one you started with.
    2. *Phase transformations* between stable *single* phases only (e.g. between austenite and ferrite in pure iron). Eventually they will be finished and no phase boundaries causing stress are left. If you start with a straight austenite blade, you may observe substantial bending during cooling but always end up with a somewhat larger straight ferrite blade.
    3. *Phase transformations* between stable *single and mixed* phases. Eventually they will be finished and lots of phase boundaries are left (e.g. between ferrite and cementite). There is a lot of stress and strain across the phase boundaries - but on large scales it averages out as long as the phases are uniformly distributed. If you start with a straight austenite blade, you may observe substantial bending during cooling but always end up with a somewhat larger and straight pearlite / ferrite or pearlite / cementite blade.
       Well - only up to a point. While you must have the same (average) carbon concentration everywhere, and thus the same amount of ferrite / cementite everywhere, the *structure* might be different. In the parts that cooled faster, cementite lamellae are closer, or you might even get bainite, compared to the regions with smaller cooling rates. That might cause some residual stress on large scales but one would not expect large effects.
    4. *Phase transformations* that produce *metastable phases*, e.g. martensite. If that happens *uniformly* everywhere in your blade, it is not different from point 3 above.
       However, if it only happens in those parts of the sample with very high cooling rates, e.g. around the edge of a blade, there will now be major residual *elastic* stress and strain across your blade, tied to the distribution of the metastable phase.
       If you start with a straight austenite blade, you observe sudden substantial bending during cooling on top of thermal gradient bending, and end up with a bend blade that contains martensite only in the edge regions and ferrite / cementite and retained austenite in a non-uniform distribution / structure in the bulk.

  The list could go on. The next point would be to look at composites, e.g. a blade with a low-carbon core and a high-carbon shell. Then we need to look at different cross sections. A bar of a homogeneous material with a *rectangular* cross-section, for example, would *not* bend because whatever happens in one colder region close to the surface and causes bending, also happens in exactly the same way on the other side, causing bending in the opposite direction - the total effect is zero! Single-edge blades, however, have a kind of triangular cross-section and therefore *will* bend.
  I will not go into this, however. The list above is good for keeping things straight in your head. It is not extremely useful in practice because for all cases where some substantial bending of a blade during cooling is expected, you will for almost sure find yourself considering the *second* major case:

- **2.** *Local stresses*, produced during cooling by any combination of the possibilities enumerated above, will *exceed a critical value* at some place and time. Something else besides just elastic deformation is then going to happen. This something else could be

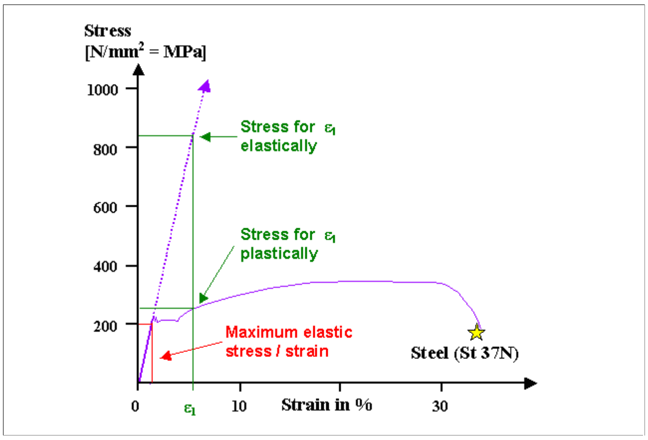
    1. *Plastic deformation* if the yield stress is exceeded.
    2. *Fracture* if some critical fracture stress is exceeded.

  Both mechanisms have something in common: They **relax**, i.e. lower the stresses, preventing that they get much larger than the critical value. That could be good but there is a price to pay:. Stress limitation leads to the destruction or at least the permanent deformation of the material. For a sword blade that could mean that it shatters into several parts, is still in one part but has cracks, or is bend but *without* containing much residual stress.

Let's look at *brittle* materials and fracture first. As far as sword blades are concerned, we don't want that. In fact we almost *never* want that.

- That is the reason why you must cool all *brittle* materials rather slowly. Large chunks of glass, just taken out of the furnace and left to cool in air, will at some point in time violently explode into thousands of pieces without prior warning. The huge glass mirrors for big telescopes must be cooled down very slowly in a process that takes many month in order to avoid self-destruction. Of all the mechanisms enumerated above only temperature gradients apply so it is conceptually easy.
  The stress relaxation mechanism is clear. As soon as the critical stress for fracture is reached, microcracks spread with the speed of sound, sending shock waves through the material. The freshly created surfaces cool down quickly, causing new temperature gradients and more crack growth - the process, once started, mushrooms into a full-fledged explosion. The pieces left over, however, will be (almost) completely stress free.

- Now let's look at *ductile* materials. In these cases you simply can never exceed the yield stress very much. This becomes clear if we look at some stress-strain curves again, slightly redrawn for the purpose here. As soon as the yield stress is reached somewhere in the sample (about 200 MPa in the figure below), plastic deformation starts and the stress just won't go up much anymore.
  - If you need to produce some strain ϵ_1 as indicated to accommodate thermal gradients or whatever, you need far less strain if the material deforms plastically than if it would only respond elastically as the figure clearly shows. The price to pay for this is that you now have plastically deformed your specimen. If you remove the stress, it will be longer, shorter, bend - whatever deformation the stress was trying to establish.



Stress needed for some strain ϵ_1 for elastic and plastic deformation

  - The shape change may be welcome or annoying - but it is far better than fracture! What exactly happens depends, of course, on the *local* value of the yield strength at any point in time. It will be different for different phases and temperatures.
  - For the same reason, whatever residual stress you might have left at room temperature because you have an inhomogeneous defect distribution (e.g. a martensite edge, a ferrite body with a little bit of pearlite, and a bainite outside for a "simple" katana) will be smaller than the local yield stress, otherwise plastic deformation would occur. In the martensite rich region of the edge the yield stress is rather large, so large residual stresses could exist in the edge of "quench-hardened" swords but not in their softer interior.
    Is this good or bad? Well - I don't know at this point but I will get back to the issue later.
- Plastic deformation during cooling is sure to happen in all sword blades that are quenched, and most likely even during normal cooling. Just consider if a piece of glass with the same dimensions would live through a similar treatment. If it shatters (and it will!) your sword blade would shatter, too, if it couldn't deform plastically.
  Now I can finally make a major point rarely considered in sword making;

Composite swords have a soft core because otherwise you couldn't make them.

- Damascene technologies of all kinds are usually only discussed in terms of the better properties they are supposed to have. "A combination of hard steel and soft iron would impart hardness and elasticity to the blade" says a major source that shall stay unnamed. B.S! I say. Whatever properties the composite will have, the first reason for using the "soft-hard" combination is simply that he blade might simply not survive its making.
  I would bet all your money that you could not make a katana without the soft core. It would simply shatter during the quenching.
  - That is a basic truth concerning most metal objects. While you, the gormless customer only care about the properties of the material, I, the engineer who has to make it, also care about the making. *You* want the body of your car to be hard but ductile. It shouldn't give easily but rebound to its old shape after you hit that tree. And even if it deforms on a major impact, it should never shatter into 1000 pieces. *I* need to deform that piece of sheet metal into a complex shape so it can be part of your car body. The softer the metal , the easier the job. The kind of metal you have in mind does exist - it just can't be shaped with ease or simply not at all.
    For your sword, as your for your car body and many other objects, we have a simple if largely unknown truth

You need to compromise between final properties and ease of making!

- I'll get back to this when I go into swords and sword making proper! But one thing is clear already now: What, exactly, is going on during the final quenching of a katana is rather complex. It is small wonder that you only get a good blade if everything is done "just right".