QUARTERLY OF APPLIED MATHEMATICS

Vol. VII JANUARY, 1950 No. 4

ELASTIC EQUILIBRIUM IN THE PRESENCE OF PERMANENT SET*

BY

G. COLONNETTI

National Research Council of Italy

Engineers, anxious to base their calculations on a rigorous as well as simple theory, do not hesitate to assume that structural materials are perfectly elastic, that is, that the deformations experienced under the action of external loads disappear as soon as these loads are removed. But this assumption is justified only if the stresses remain everywhere below a certain limit, called the "elastic limit", and if the period over which the loads are applied is sufficiently short.

In practice, the stresses exceed the elastic limit much more frequently than is generally supposed, even in the most rigorously and safely calculated structures. The duration of the stresses, on the other hand, is extremely variable, covering all the range from impact stresses to stresses which subsist during the entire lifetime of the structure.

Stresses beyond the elastic limit, as well as long sustained stresses, produce non-elastic strains which do not disappear on removal of the loads which have caused them. This fact has two important consequences. First, the assumption of a one-to-one correspondence between strains and stresses, which is fundamental in the theory of elasticity, must be abandoned. Secondly, a state of residual stresses** is established, causing a distribution of the inner stresses which is completely different from the distribution furnished by the theory of elasticity.

Obviously, we cannot represent this phenomenon by a mathematical theory without introducing some hypothesis regarding the nature of the deformation and its relations to the stresses which produce it. This is what we are proposing to do next, and we will try to find the physical significance of these hypotheses and point out to what extent they correspond to the empirical facts.

A solid body is said to be perfectly elastic if there exists in it a one-to-one correspondence between stresses and strains. This is the case for most structural materials, as long as the stresses are sufficiently small and act for a sufficiently short time. But, as soon as the stresses become greater, though they may act over a very short period, the process ceases to be reversible; part of the strain will remain in the structure even after the loads have been removed. We call these "plastic strains".

Of course, we do not intend to assert that there actually exists, for each material, a well-determined limit such that the material behaves in a perfectly elastic manner

---

*Received April 4, 1949. This paper was presented at the Second Symposium on Plasticity sponsored jointly by the Office of Naval Research and the Bureau of Ships (Brown University, April 4-6, 1949).

**The term "residual stresses" as used in the following refers to the state of stress as well as to the state of permanent strain which causes the residual stresses. It would be desirable to adopt a single English term for this concept corresponding to the Italian *coazione* or the French *coaction*.

as long as the stresses do not exceed this limit. We wish to state only that, for all practical purposes, we may assign a limit, below which the process can be treated as reversible, in the sense that the plastic strains are negligible in comparison with the elastic strains. Beyond this limit, however, the process becomes decidedly irreversible. In general, this limit is purely conventional. As a matter of fact, the limit, beyond which irreversible processes are observed, decreases with increasing accuracy of the observations.

However, there are materials, such as iron, mild steel, bronze, etc., for which the limit between the elastic range (characterized by a nearly complete absence of permanent deformation) and plastic range (where almost the whole deformation is permanent) is well-defined. In these materials the permanent strains, though occurring also when the stresses are very small, are negligible compared with the elastic strains as long as the stresses do not exceed a certain critical value. As soon as this value is reached, the laws which govern the process change all of a sudden, and we may say that all further strains are practically permanent. As a matter of fact, the plastic strains which now occur, grow so rapidly, even under a constant stress, that the elastic strains which affected the material up to this moment, become negligible in comparison. This limit, which is no longer merely conventional but has a well-determined physical meaning, is called the "yield limit".

We will choose this apparently rather special case as the starting point to derive a mathematical theory of the elasto-plastic equilibrium, that is of the equilibrium of bodies in which the elastic limit has been reached and plastic deformations have occurred.

We will make the following assumptions: first, the one-to-one correspondence between stresses and strains is valid for each element of the body as long as the material is stressed within its elastic range; secondly, as soon as this limit has been reached, the element becomes perfectly plastic and the deformation can increase indefinitely under constant stresses. Moreover, we will assume that the plastic strains are superimposed on the elastic strains without changing the characteristics of the latter, particularly without depriving them of the property of disappearing upon the removal of the loads which have caused them.

It can be shown that these apparently very restrictive hypotheses adequately cover a great number of cases. They even explain the cases in which the transition from the elastic zone to the plastic zone takes place with continuity, that is, when a physically well-determined elastic limit is absent. From these hypotheses it is also possible to develop a very elegant interpretation of the so-called "strain-hardening", which is always connected with plastic deformation.

To this end we will first investigate the nature and the fundamental characteristics of this phenomenon, referring to a very particular, but simple and experimentally well-known case: that of a cylindrical or prismatical metallic bar submitted to axial tension. It is well-known that when the yield limit is reached in such a bar, localized slip occurs along certain planes, whose angle with the axis of the bar is determined by a relation between the normal and the shearing components of the stress transmitted across these planes.

A single, small inhomogeneity in the material is sufficient to start localized slip. In the zones where slip occurs it has the tendency to cause a reduction of the cross section together with an increased strength of the material. At first, the favourable effect of this second factor more than balances the unfavourable consequences of the first one. Therefore, slip soon ceases in that section only to appear immediately in some other section, then in a third one, and so on. This process of propagation may be made visible by polishing carefully the surface of the test specimen. Slip is then indicated by the appearance on the polished surface of extremely thin inclined lines, the so-called slip bands. These bands grow in number, extend, and become interlaced in a kind of opaque veil which quickly spreads and soon covers the whole surface uniformly.

We are thus led to consider a new condition of the material in which the original homogeneity, obviously destroyed by localized slip, is finally restored on the whole. The most remarkable and obvious characteristic of this new condition is a rise of the elastic limit of the material, that is, an increase of the value of the stresses to which the test specimen can be submitted without showing plastic strains. It is exactly this rise of the elastic limit which is called the strain-hardening of the material.

Having reached these conclusions from the experimental standpoint, we are now going to look at the problem from an exclusively theoretical point of view. We will assume that the material is perfectly homogeneous and consequently that in the elastic field the stresses are uniformly distributed throughout the specimen. According to this hypothesis, the elastic limit will be reached in the same instant in all parts of the body, and we may conclude that the plastic strains will be uniformly distributed too.

These plastic strains will not give rise to any residual stresses; they will be superimposed on the elastic strains, without changing the stress distribution or the value of the strain energy. This means that the mechanical work necessary to produce these plastic strains has been completely spent to overcome the inner friction and to produce heat.

Let us now suppose that the specimen is unloaded. The conclusions reached above explain the fact that the specimen has a permanent set, in the strictly geometrical sense of a permanent change of shape and dimensions, but they do not indicate at all the experimentally established raising of the elastic limit.

However, due consideration must be given to the fact that the assumed structural homogeneity of the material, as well as the uniform distribution of the plastic strains, do not really occur. It then becomes clear that, since the plastic strains are no longer "compatible" as a rule, a state of residual stresses is set up, which will superimpose itself on the stresses previously established in the elastic range and which will subsist even after all elastic strains have disappeared. In order that this condition may be achieved, at least a portion of the mechanical work spent on producing the elastic strains should not be lost, but transformed into potential energy. This energy will not be recovered as external work when the loads are removed. For this reason we shall call it "residual energy".

However, the simple fact that this residual energy subsists after the external forces have ceased acting, is sufficient to let us think that something else beyond shape and dimension has changed in the body, and that here lies, perhaps, the explanation of the phenomenon with which we are dealing. We will show that this is possible by means of a very simple but very significant example, taken from the field of reinforced concrete design.

Let us consider a reinforced concrete bar submitted to an axial tension. It is well-known that, as long as the stresses are small and the material behaves everywhere in a perfectly elastic manner, the stresses must be distributed over the various elements of each cross section of the bar according to the ratios of their moduli of elasticity, if the cross section is to remain plane. But as soon as the elastic limit of concrete is reached, the material enters the plastic state. Again, if the cross section is to remain plane, the concrete must undergo, independently from any further increase of the stresses, such strains as are necessary to let the metallic reinforcement carry any further increase of the load.

The stress-strain diagram then takes the shape of a broken line, of which the first section, beginning at the origin, covers the elastic range. The second section covers the elastic-plastic range. A third section, parallel to the strains axis will then represent the final phase, when the entire cross section behaves in a plastic manner.

Let us now suppose that the load is removed. The elastic strains will naturally show a tendency to disappear; this will not be the case for plastic strains. On the contrary, these will even succeed in preventing a complete disappearance of the elastic strains. Thus, a state of residual stress will be set up, such that the metallic reinforcement will be stressed in tension, while the concrete will be under compressive stresses. It follows from this that the load required in order to let the concrete again reach its elastic limit, so that further plastic strains occur, must now be greater than the first time.

If we ignore the existence of the residual stresses, we shall naturally speak of a rise of the tensile elastic limit of the concrete. But this is not correct: it is not the elastic limit of the concrete which has risen, it is the initial state of stress which is no longer the same.

This result is not at all restricted to the type of structure we have considered as an example. To make sure of that, let us go back to the fundamental problem, from which we started, and study the general case of any structure whatsoever, consisting of heterogeneous elements having elastic limits varying from a certain minimum to some maximum. Assuming these elements to be uniformly distributed throughout the structure, we can see in a very general way that the above-mentioned broken line will be replaced by a polygonal line and, as a limit, by a curve. In fact, each element stops contributing to the resistance of the structure as soon as it reaches its own elastic limit, and thereafter it just conforms plastically to the elastic strains of the surrounding elements, which have higher elastic limits. These elements must carry any further increase of the load. In other terms, the stress is shifted from the elements which have reached their yield limit to the elements which are still stressed elastically. If now the loads are removed, the plastic strains prevent the structure from returning to the original unstrained shape and thus give rise to residual stresses. The elements which, owing to their higher yield limit, have carried the greatest portion of the load, keep on being subject to stresses of the same type, whereas the elements which have undergone plastic strains show a tendency to be subject to stresses of the opposite type.

It follows from this that when the structure is again loaded, the last elements are in a more favourable condition, because the residual stresses will have to be overcome before stresses of the type connected with the acting loads may again appear. Thus the elastic limit shall be reached, even for the weaker elements, when the loads have attained a higher value than was originally necessary to produce plastic strain. This explains the apparent rise of the elastic limit of the material.

If, on the other hand, after having produced the strain hardening of a specimen, by submitting it to the action of a well-determined system of loads, we apply to the same specimen a system of loads of the opposite direction, the stresses produced by these forces will be added to, and not subtracted from, the residual stresses. It will then appear that the elastic limit has been lowered (Bauschinger effect).

We are thus entitled to consider our interpretation of the phenomenon of strain-hardening as reasonable. At any rate, it is certain that the outline of the phenomenon, which we have suggested as a starting point of a mathematical theory of the elasto-plastic equilibrium, leads to a very satisfactory interpretation of the experimental results.

The situation is very different when the stresses, though not being high enough to cause the elastic limit to be reached in any point of the structure, keep on acting for a time long enough to result in irreversible effects, that is to give rise to strains which are not going to disappear immediately when the loads producing them are removed. This phenomenon has been long known as "time effect" and involves the superposition on the elastic strains, due to the external forces, of other strains which grow progressively with time under the sustained action of those same forces. These strains do not disappear as soon as the forces cease acting, but they may disappear later on (in any case very slowly) if the structure remains free to return to its initial unstrained shape.

A mathematical theory of such a phenomenon is not easy to establish, because the strains at a certain instant depend not only on the loads acting at this instant, but also on previous loads and on the duration of the period for which they have acted. We shall consider only a very particular case, which is, however, very important for practical purposes, namely, the case of very small strains (of the magnitude of the elastic strains) which we shall assume to increase, at any instant, in proportion with the stress existing at this instant.

Let us suppose that the structure is homogeneous and that the coefficient of proportionality, which is usually called the "creep coefficient", is the same for the entire structure. Moreover, let us assume that there is no state of initial stress. The time effect will then involve strains which, like the elastic strains, satisfy the conditions of compatibility. Accordingly, these strains will subsist even when the elastic strains will have disappeared with the removal of the loads. It is clear, therefore, that the time effect does not produce any residual stresses, if the structure was free of initial stresses. It can be shown quite generally that in a homogeneous body the time effect appears geometrically as a proportional increase of the elastic deformation produced by the loads and statically as a proportional decrease of the residual stresses.

Let us now introduce the hypothesis that the body is not homogeneous, which proved so useful when we dealt with plastic strains. The situation will then change, and the theory will explain immediately the observed tendency of the strains, due to time effect, to vanish slowly once the loads have been removed.

We will assume, therefore, that the body consists of heterogeneous elements affected by different flow coefficients, uniformly distributed throughout the body. We may foresee that the elements characterized by higher flow coefficients, following a prolonged action of the external forces, will undergo greater strains than the elements with lower values of that coefficient. As a consequence, the last elements, conforming to the first, will undergo supplementary elastic strains. When the loads are removed, these additional strains, like every elastic deformation, will tend to vanish, but they will be prevented from vanishing completely by the non-elastic strains of the surrounding elements. This gives rise to a state of residual stress: the stresses caused by the elements with small non-elastic strains will produce in the elements with greater inelastic strains a new set of strains of opposite type, and these will force the body to move towards the initial unstrained shape.

From this mass of apparently so different and complicated facts a very interesting and absolutely general conclusion can be drawn concerning the co-existence, at any moment and in any state of equilibrium, of elastic strains and non-elastic strains, whatever their origin and their physical nature may be. Here is how we can reach this conclusion quickly.

We will denote the components of the elastic strain by

$$\epsilon_x\,,\ \epsilon_y\,,\ \cdots\,,\ \gamma_{xy}$$

and those of the plastic strain by

$$\bar{\epsilon}_x\,,\ \bar{\epsilon}_y\,,\ \cdots\,,\ \bar{\gamma}_{xy}\,.$$

We will further assume that both strain distributions are continuous, uniform and very small, so that it is always possible to write the components of the resulting total strain as follows:

$$\epsilon_x + \bar{\epsilon}_x\,, \qquad \epsilon_y + \bar{\epsilon}_y\,, \qquad \cdots\,, \qquad \gamma_{xy} + \bar{\gamma}_{xy}\,.$$

It must be kept in mind that the resulting total strains are compatible, but that neither one of the two partial strains will be compatible as a rule. Assuming that neither one of the two partial strains is compatible, let us now try to state analytically that one of these partial strains is elastic.

This can be achieved in two ways, either considering the deformation as a state of equilibrium, or taking into account the way it occurs or vanishes.

In the first case we are led to assume that any element of the body is kept in a strained condition by the stresses which act on it owing to the surrounding elements. Thus the strains are strictly connected with a system of stresses which we shall define, as usual, by means of six components of stress, namely:

$$\sigma_x\,,\ \sigma_y\,,\ \cdots\,,\ \tau_{xy}\,.$$

If, on the other hand, we adopt the second point of view, taking into account the tendency of the strains to vanish once the loads which caused them have been removed, we must conclude that there exists an elastic strain energy, produced by the transformation of the mechanical work spent to make the actual strains occur in the body. This energy may again be recovered as external work, when the strains disappear for one reason or the other. This strain energy is assumed to be the sum of the elementary strain energies of all the individual elements of the body. In this way each of these elementary energies is completely defined when the condition of the element to which it refers is known. It is not necessary, in order to characterize it, to introduce either the condition of the other elements or their relative position. This results in the following expression of the energy of the body:

$$\Phi = \int_V \varphi\, dV,$$

where φ is the elastic strain energy referred to the unit of volume. By definition, φ is a function of the six strain components.

It can then be shown that the elementary elastic strain energy is a positive definite form of the second order, whose first partial derivatives with respect to the six strain components equal the corresponding stress components. Thus the assumption from which we started amounts to this: the six stress components are linear and homogeneous functions of the six strain components, and vice versa. It follows from this that the elementary elastic strain energy may also be regarded as a second-order, positive definite form of the six stress components. Its first partial derivatives with respect to these new variables equal the corresponding strain components.

Let us now study the non-elastic strains, that is, the strains which do not vanish together with the loads. The work spent in order to produce these strains will not be recovered immediately. To develop a better understanding of this we will study separately the element, taken alone, and the body consisting of all elements.

In the first case, the fact that the mechanical work spent to produce the strains cannot be recovered may be expressed by assuming that this work is lost through irreversible transformations. This prevents the accumulation of any strain energy. The same idea can equally be expressed by stating that the strains, once they have been produced, go on existing independently from any external action which may keep the body in its strained condition. As a consequence, the existence of strains does not imply the presence of stresses.

But what has been stated for a single element does not apply to the body as a whole. In it non-elastic strains can occur only in some points, so that they will generally form a non-compatible system. Along with this system we must then have a series of complementary elastic strains such that the composite strains become compatible. Thus the non-elastic strain of a given element gives rise to a series of elastic strains and of stresses in the surrounding elements. We shall call these stresses "residual stresses" to distinguish them from the stresses which are connected with the loads. These stresses involve an elastic energy which we shall call "residual energy" to remind us that it is connected, like the condition to which it refers, with the presence of non-elastic strains and that it keeps on existing even after the loads which produced it.

The mechanical work required by a non-elastic deformation does then consist of two parts: one is spent to produce the deformation itself and is lost through irreversible processes; the other gives rise to the state of residual stress. Neither of these two functions is recovered the moment the loads are removed. But while the first one is definitely lost, the second will remain available in the body as strain energy. It may even happen that, joining the strain energy produced in the same body by another system of loads, this residual energy will become apparent through the reactions of the body. To the new loads, it may even be transformed again into mechanical work, if there exist in the body any elements able to be further strained under time effect, or if by proper treatment, we succeed in freeing it. In any case, this distinction between the two parts of the work spent on the production of non-elastic strains is of basic importance. It will always be necessary to take into account separately the part which causes a real loss of energy, while, in view of global evaluation of the strain energy, the other fraction will have to be treated as if it were work spent in order to produce new elastic strains.

Let us now consider the body in equilibrium under the action of a given system of loads and suppose that the stresses undergo a very small variation compatible with these loads. This means that, while the non-elastic strains remain unaltered, the stress components shall undergo very small variations

$$\delta\sigma_x\ ,\ \delta\sigma_y\ ,\ \cdots\ ,\ \delta\tau_{xy}\ ,$$

which form a system of stresses in equilibrium in the absence of external forces.

We consider the function

$$\Phi + \int (\bar{\epsilon}_x \sigma_x + \cdots) \, dV,$$

the first variation of which can be written in the form:

$$\int_V [(\epsilon_x + \bar{\epsilon}_x) \, \delta\sigma_x + (\epsilon_y + \bar{\epsilon}_y) \, \delta\sigma_y + \cdots + (\gamma_{xy} + \bar{\gamma}_{xy}) \, \delta\tau_{xy}] \, dV.$$

We see at once that this variation yields the work which the system of stresses

$$\delta\sigma_x \, , \; \delta\sigma_y \, , \; \cdots \, , \; \delta\tau_{xy}$$

would carry out if the body were to undergo the change of shape characterized by the following components:

$$\epsilon_x + \bar{\epsilon}_x \, , \qquad \epsilon_y + \bar{\epsilon}_y \, , \qquad \cdots \, , \qquad \gamma_{xy} + \bar{\gamma}_{xy} \, .$$

Now this system of stresses is, by hypothesis, in equilibrium in the absence of all external forces. On the other hand, the change of shape is certainly compatible with the possible constraints of the body, since it is exactly the one which the body had to undergo in order to pass from the unstrained natural state on to the state of equilibrium which we are considering. From the principle of virtual works it follows, therefore, that

$$\int_V [(\epsilon_x + \bar{\epsilon}_x) \, \delta\sigma_x + \cdots] \, dV = 0.$$

Taking into account the fact that the second variation is:

$$\int_V \varphi(\delta\sigma_x \cdots) \, dV$$

and that the integrand is positive definite, we may state the following theorem: *For each system of loads and of non-elastic strains, the inner stresses characterizing the state of equilibrium are those which minimize the expression*

$$\Phi + \int_V (\bar{\epsilon}_x \sigma_x + \bar{\epsilon}_y \sigma_y + \cdots + \bar{\gamma}_{xy} \tau_{xy}) \, dV.$$

If there are no non-elastic strains, the problem fits into the framework of the classical theory of elasticity, and the function which has to be made a minimum is the elastic strain energy. We find here again, as it could have been foreseen, the theorem of Menabrea.

It is of basic importance to note that we can pass from the equation which is the basis of the classical theory of elasticity, to the equation which appears to be a possible basis of a new theory of the elasto-plastic equilibrium, simply by replacing the six elastic strain components by the six components of the total strain. Provided an analogy exists between the surface conditions involved in each particular problem, the solution to which the classical theory of elasticity leads for its problems may, therefore, be transformed into as many solutions of analogous problems of elasto-plastic equilibrium.

In this way we are able to establish a new theory of the strength of beams, the most important conclusion of which is a new outlook concerning the redundancy of constraints and the consequent statical indeterminacy of the structures. By the classical theory of elasticity one has become accustomed to consider statically indeterminate structures as extremely sensible to all non-elastic strains. This sensibility was often considered as a disadvantage, sometimes even as a real danger, since non-elastic strains were beyond the scope of our calculations and, consequently, of our forecasts. Today, as this scope has extended to cover the whole field of non-elastic strains and we are able to take exact account of the effects of these strains and to calculate the changes they will cause in the state of equilibrium of the structures, we should consider this sensibility rather as an advantageous and even precious quality of the hyperstatic structures.

Let us consider, for instance, the case of proper plastic strains, that is, the strains which occur in the outermost fibers of a beam as soon as the elastic limit of the material has been reached. Our theory points out in the most simple and clear way three types of effects which such strains may have. The first type concerns exclusively the distribution of stresses over the section itself and becomes apparent by limiting the stresses in the fibers where the elastic limit has been reached: this limitation is compensated for by an increase of stresses in the inner zone of the section, where the strength of the material was less utilized originally. This new stress distribution is due to the static indeterminacy of the cross-section itself; it takes place, therefore, even when the beam is statically determinate, as far as external constraints are concerned. The second type, on the other hand, involves the static indeterminacy of the diagram of bending moments, that is, the existence of at least one redundant constraint; it becomes apparent by limiting the value of the moment in the section where the plastic strains have occurred and by increasing it in the sections which were less stressed originally. The third type, finally, is observed only if the constraints involve further static indeterminacy of the structure; it becomes apparent through a force which modifies the distribution of the stresses not in one or more sections, where plastic strains have occurred, but in all the sections of the beam. This force tends to limit that, of the two maximum stresses, which first has reached the elastic limit, and to increase the other.

It is by a proper combination of these effects, that is, by using each of them in the most suitable way and by following laws which only a rigorous and complete analysis of the phenomenon can discover, that nature carries out, within the limits fixed by the data of the problem, the best possible utilization of the resistance of the material. Besides, it is clear that this behaviour of the structure, which is now at our disposal since we know its secret, will have a wider field of application and will allow us to obtain important simplifications of the analysis of stresses in indeterminate structures. Therefore, we must get used to looking at the problem in a new way.

The classical theory had led us to think of the state of equilibrium of a system submitted to given loads as something well-determined in terms of these loads. Even in the statically indeterminate case, in which the equation of statics yields an infinite number of solutions, the hypothesis of perfect elasticity of the material is sufficient to determine the solution completely. This result does not represent the truth, however. Whenever elastic strains are accompanied by even very small non-elastic strains, it is no longer true that only one solution of the equations of statics is physically possible. On the contrary, they all are possible, and which one shall come into being depends upon the occurring of suitable non-elastic strains and of the state of residual stress connected with them.

But here the field of application widens. The question is no longer about the plastic strains only and about the part they can play in determining the state of equilibrium.

As a matter of fact, the study of plastic strains has allowed us to catch a glimpse of possibilities which can be accomplished in quite a different way. It is sufficient to recall the modern technique of the prestressed materials, especially of prestressed concrete, to get an idea of the wide applications of the theorem which we have just derived and of its use in the study of the strength of materials.

Acknowledgement: The author thanks Ing. Pier Luigi Romita, who has prepared the English text of this paper.