THE SOLAR CYCLE - ITS INTERPRETATION IN TERMS OF A DYNAMO THEORY

Hirokazu Yoshimura
Department of Astronomy, University of Tokyo

First of all, I must explicitly state that this is not a review talk in usual sense. I concentrate on the progresses that I have followed myself over these 10 years. Some of you may feel the theory and the interpretation somewhat exotic and unfamiliar. As a matter of fact, however, the theory I will present here is a natural consequence of world-wide advancement of understanding of the solar cycle.

1. Observation

The phenomena of the solar cycle comprise every aspect of the solar activity and its related phenomena which show a definite cyclic behavior with 11 years period. Since the solar activity is a magnetic activity, the solar cycle represents an oscillating state of the magnetic fields in the interior of the Sun. The cycle period is doubled to 22 years when we take into account polarity of the fields. We shall consider here briefly the following questions:

(i) why are the magnetic fields created in the Sun?
(ii) why are the fields in an oscillating state?
(iii) how are they related to individual phenomena of the solar activity?
(iv) how does the solar cycle vary over longer time scales and why?
(v) what do the phenomena imply for studies of solar-like late type stars as well as stars of general categories?

Let me first remind you what the solar cycle is, and what characteristics of the phenomena we will consider to be basic for our understanding the phenomena. First, we know the time series of the sunspot relative number. This historical compilation of data is basic to our knowledge of the Sun and other cosmical objects in general. It tells us that the Sun, and a star in general, can be in a magnetically oscillating state. It tells us that the oscillating state can undergo various kinds of long-term modulations like the 80- or 55-year grand cycles or the series of Maunder Minima. Historical compilation of naked eye sunspots, as well as ^{14}C data deposited in organic material on the Earth, as we will discuss later, suggest that the Sun is undergoing modulations of much longer time scales, like hundreds and thousands years. For its simplicity and easiness to obtain, the sunspot relative number should remain as an important index to register the long-term behavior of the Sun. Other indices so far devised, are either insensitive to the oscillating nature of the Sun or contaminated by other rather independent factors.

Secondly, we know the time evolution of surface distribution of sunspots and other solar active phenomena. We know the Maunder Butterfly Diagram of sunspots and the Diagram of Prominences by L. and M. d'Azambuja and by Ananthakrishnan. These diagrams show evolutionary structure of latitudinal distribution of the phenomena. When combined with its intensity and polarity and displayed by contour plots, the diagrams represent the following basic characteristics of the solar cycle:

(i) Schwabe-Wolf's law of the existence of the periodic oscillation of the sunspot occurrence,

(ii) Carrington-Spörer's law of the equatorial drift of the sunspot zones,

(iii) Hale-Nicholson's polarity rules of alternations in magnetic-field polarity of bipolar sunspot groups for the northern and southern hemispheres and for consecutive cycles,

(iv) the poleward migration of polar prominences,

(v) the slight indication that the sunspot zones also migrate towards the poles,

(vi) the cyclic variation of the number of higher latitude faculae and the phase shift between it and the lower latitude sunspot number variation.

In relation to the phenomena at the poles, we have

(vii) Babcock's law of polarity reversals of the solar polar fields.

Thirdly, we know the solar-cycle-associated changes of the corona and the related phenomena in the interplanetary space and the heliosphere. We know

(viii) the round shape of the corona at a maximum phase and its typical dipole-like shape with long streamers at the equator at a minimum phase of the solar cycle,

(ix) the equatorial migration of the low latitude coronal activity and the poleward migration of the high latitude activity as seen in 5303 Å lines,

(x) the Forbush's negative correlation between the solar cycle and the galactic cosmic rays,

(xi) the negative correlation between the solar cycle and the high speed solar wind.

Fourthly, we come to know quite recently that

(xii) the rotational flows of the Sun also undergo solar-cycle-associated torsional oscillations.

According to Howard and LaBonte, the amplitude of the oscillations is on the order of 3m/sec. The wave pattern of the oscillations is symmetric about the equator and consists of four alternating zones of fast and slow rotation in a hemisphere. The whole pattern propagates from the poles to the equator with time scale of about 22 years. The latitude of maximum activity corresponds to the poleward boundary of a fast zone in low latitude zones.

These are the characteristics which have been known before the theory was developed. All of these are characteristics associated with one 22-year solar cycle. We will discuss later other characteristics associated with the long-term behavior of the solar cycle.

2. Theory

We now turn our attention to the theoretical aspect of the problem. In the development of understanding of the phenomena, we see a typical interplay of theory and observation. Theory is guided by hypotheses and their deduced consequences. Sometimes we see jumps of logic guided by intuition. These hypothesis and intuition will be verified later by observation. In fact, the theory of the solar cycle not only has given explanation and interpretation of the already known phenomena but also has predicted unnoticed characteristics of the phenomena which were later confirmed by analysing the observed data. The theory and the observation are progressing now side by side. Sometimes observation precedes theory and sometimes theory precedes observation.

We start from the governing equations; the well known equation of motions, the equation of continuity, the equation of energy transfer, and the MHD induction equation and the equation of magnetic flux conservation. If we can solve this set of equations for any general situations, any phenomena in astrophysics, where continuum approximation is valid, will be reproduced. However, this is impossible, at least at present, especially because the phenomena as we see in nature contain a multitude of hierarchy of scales and so great variety of components. In such circumstances, we filter out the equations and simplify them using various kinds of approximations so that the equations are suitable for representing a phenomenon or a set of phenomena in mind, which are generally conceived by observing the nature. We do not ask whether the observation is correct or precise. Just a concept is enough at this stage.

2.1) The Concept of the Global Convection

In 1960's, a concept has emerged from the magnetograph observation of global distribution of magnetic fields at Mt. Wilson and from the statistics of sunspot proper motions of Greenwich Photoheliographic Results, that there seem to be giant scale flows in the Sun. Howard, Bumba, and Smith at Mt. Wilson thought it was convection, like the supergranulation, which was just discovered at those times. Ward, who analysed the sunspot proper motions, thought it was a kind of Rossby waves just like the general circulation in the Earth's atmosphere. The two kinds of flows correspond to the two of the three possible modes of motions in a rotating fluid sphere. The other is the sound wave modes with much shorter time scales. The difference between the former two is that vertical motions are important for the convective or gravity modes while horizontal motions dominate in the Rossby modes. Hence vertical heating is efficient to drive the convective modes while horizontal heating can easily excite the Rossby modes. Since no appreciable latitudinal temperature difference between the poles and the equator has been observed, the hypothesis of the Rossby waves in the Sun has eventually been dropped out. The effects of rotation on these two are also greatly different. So, we consider the convective modes hereafter and call it the global convection.

2.2) Linear Studies of the Nonzonal Velocity and Magnetic Fields

The first step is to study its linear structure and solve the linear set of equations for the velocity as well as magnetic fields. We study especially the effects of rotation on the field pattern. Even at this stage, it is known that the following important characteristics of solar activity can be understood if the global convection really exists in the solar convection zone:

(i) the existence of Unipolar Magnetic Regions (UMR's) and ghost UMR's and their rigid-body like rotation,

(ii) the existence of active longitudes and complexes of activity and their rigid-body like rotation,

(iii) the faster (slower) rotation of sunspots (magnetic features) than the averaged rotation of the solar mass in low (high) latitudes,

(iv) the preponderance of preceding spots of bipolar sunspot groups to the following spots,

(v) the tilt of bipolar axes of the sunspot groups,

(vi) the forward inclination of normal axes of sunspots inferred from the east-west asymmetry of the appearance and total area of sunspots,

(vii) the association of the characteristics of an active region with the presence of an older active region in its vicinity and with the relative disposition of the two active regions; that is, according to Martres a sunspot group formed in the preceding sides of an older AR evolves to the type βf, the following spot showing a greater eastward motion than in the case of an isolated group, while a sunspot group in the following side of an older AR evolves to the type βp which is associated with a greater westward motion of the leading spot.

2.3) Formulation of the Equations governing the Zonal Fields

Second step is to study the second-order nonlinear effects of the flows. To study the effects, we perform another scaling. That is, we separate the fields into zonal fields <Q> and nonzonal fields Q'. The differential rotation and the general magnetic fields of the solar cycle are the zonal fields averaged over longitude. Nonzonal fields are fields that are associated with the global convection. We construct equations governing the zonal fields which contain the nonlinear effects of the nonzonal fields. Second-order correlation of the nonzonal velocity fields themselves shows that the global convection can transport angular momentum latitudinally and accelerate the equator preferentially. This problem depends on the modes of the convection which are achieved in the Sun and no final solution is obtained yet. Second-order correlation of the velocity and magnetic fields, which appears in the equation governing the zonal magnetic fields and which is deduced from the induction equation, gives rise to the growth of the magnetic field from an infinitesimal level of the field. In other words, the global convection can have the dynamo action. The governing equations thus obtained are as follows:

$$\frac{\partial \langle \underset{\sim}{H} \rangle}{\partial t} = \underset{\sim}{\nabla} \times (\langle \underset{\sim}{V} \rangle \times \langle \underset{\sim}{H} \rangle) + \underset{\sim}{\nabla} \times \langle (\underset{\sim}{V}' \times \underset{\sim}{H}') \rangle - \eta \underset{\sim}{\nabla} \times \underset{\sim}{\nabla} \times \langle \underset{\sim}{H} \rangle \quad , \qquad \text{(Eq.1)}$$

$$\underset{\sim}{\nabla} \cdot \underset{\sim}{H} = 0 \quad , \qquad \text{(Eq.2)}$$

which can be reduced to

$$\frac{\partial \Psi}{\partial t} = R\,\Phi + \left(\frac{(1-\mu^2)}{r^2}\frac{\partial^2}{\partial \mu^2} + \frac{\partial^2}{\partial r^2}\right)\Psi \quad , \qquad \text{(Eq.3)}$$

$$\frac{\partial \Phi}{\partial t} = G\,\Psi + \left(\frac{(1-\mu^2)}{r^2}\frac{\partial^2}{\partial \mu^2} + \frac{\partial^2}{\partial r^2}\right)\Phi \quad , \qquad \text{(Eq.4)}$$

where

$$\Psi = A\, r \cos\theta \quad , \qquad \Phi = B\, r \cos\theta \quad , \qquad \text{(Eq.5)}$$

$$\langle H_\phi \rangle = B \ , \ \langle H_\theta \rangle = \frac{1}{r}\frac{\partial}{\partial r}\, r\, A \ , \ \langle H_r \rangle = -\frac{1}{r\cos\theta}\frac{\partial}{\partial\theta} \cos\theta\, A \ , \qquad \text{(Eq.6)}$$

and

$$G = (1 - \mu^2)\left(\frac{\partial\Omega}{\partial\mu}\frac{\partial}{\partial r} - \frac{\partial\Omega}{\partial r}\frac{\partial}{\partial\mu}\right) \ , \qquad \text{(Eq.7)}$$

$$V_\phi = \Omega\, r \cos\theta \quad , \qquad \text{(Eq.8)}$$

and R is the regeneration factor. We normalize G and R by N_G and N_R, called the dynamo numbers representing the strength of the generation and regeneration processes. These equations are characterized by Ω and R.

Here we adopt a jump of logic and generalize the results obtained by the linear solutions using various approximations. We deform the profile of R topologically. If R is to be deduced from a certain type of flows, those flows can only be achieved in compressible stratified systems with proper effects of small scale turbulence.

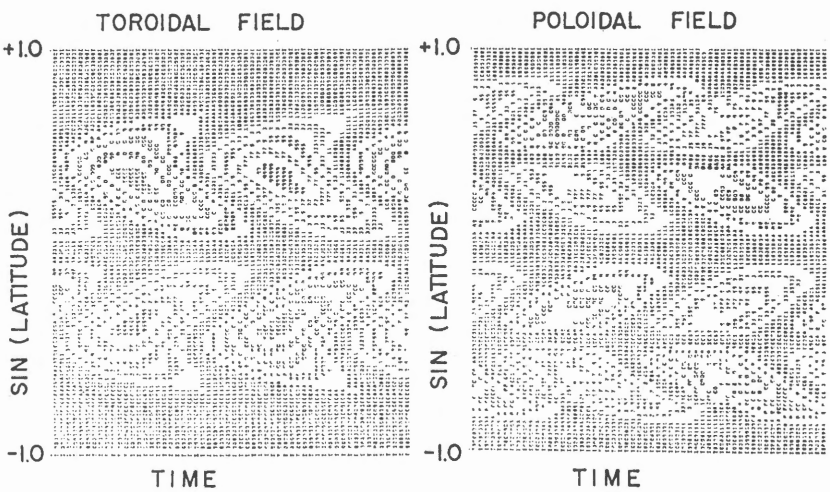

Fig. 1. The solar-cycle-simulating surface patterns of the solutions of the dynamo equation. The toroidal field pattern corresponds to the Butterfly diagram of sunspots. The poloidal field pattern corresponds to the Butterfly diagram of the surface poloidal general fields of Fig. 2.

In other words, a particular profile of R contains such various kinds of dynamics.

2.4) Linear Dynamo Problem and Characteristics of One Solar Cycle

Next step is to solve this set of equations, called the dynamo equations, for a given set of Ω and R. We solve them by a numerical method by computers. An example is shown in Figure 1. The equations up to this point are linear with respect to the general magnetic fields $\langle \underset{\sim}{H} \rangle$. Solutions of such equations become oscillatory waves which we call the dynamo waves. The characteristics of the dynamo waves as simulated by the equations can reproduce and explain the observed characteristics of the solar cycle described above. Growth of the solar magnetic fields and their oscillatory nature are thus attributed to the growth and oscillation of the dynamo waves driven by the differential rotation and the global convection. An interesting and basic property of the dynamo waves is that the waves propagate along isorotation surface of the differential rotation Ω. Since different set of Ω and R gives different set of surface pattern, we can infer the structure of Ω and R that are achieved in the Sun by choosing such set of Ω and R that give rise to the patterns which are similar to the observed ones.

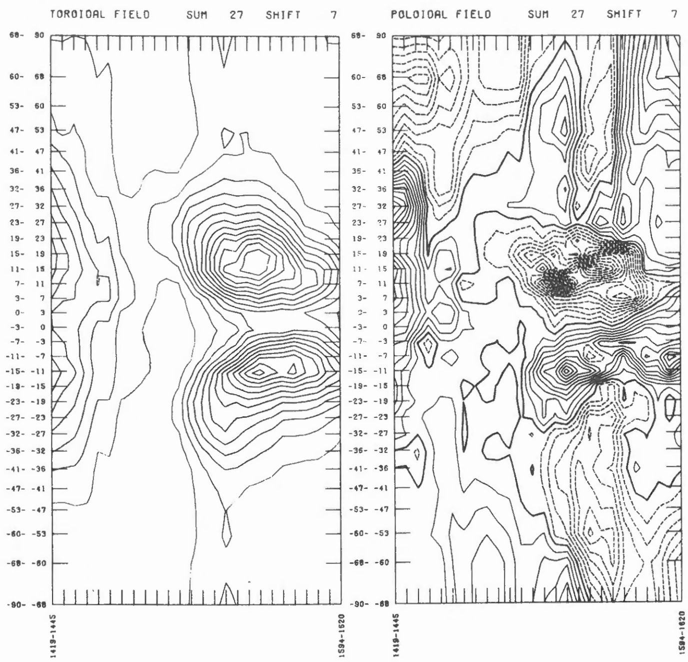

Fig. 2. The Butterfly Diagrams of the poloidal general magnetic fields (right) and of the toroidal general magnetic fields (left) deduced from the data observed at the surface.

At this stage, I myself encountered with a difficulty. The observed characteristics at those times could not determine how strongly the latitudinal gradient dominated the differential rotation. If the radial gradient exists, even slightly, the dynamo waves migrate towards the equator as the sunspot zones do. I needed an independent information to determine the degree of importance of the latitudinal gradient. The evolutionary surface pattern of the poloidal field, or of the radial component of the fields, could discern the cases with and without strong latitudinal shear. Especially the behavior of the fields at high latitudes was the key factor. So I analysed the data of the synoptic magnetic data, observed and accumulated at Mt. Wilson, and obtained the Butterfly Diagram of the poloidal fields as displayed in Figure 2. This diagram clearly shows that the latitudinal shear is important for the formation of the polar branches of the poloidal fields. At the same time, this diagram combines the various observations concerning the solar cycle phenomena at the polar region as listed above. This diagram gives a physical basis to the understanding of the prominence diagram of d'Azambujas and Ananthakrishnan. Although the latter does not have information concerning the polarity of the magnetic field, it does show that there certainly exist the polar magnetic branches migrating towards the poles.

Then I calculated the coronal magnetic fields associated with the solar cycle poloidal fields at the surface, obtained by theory as well as by observation. This study gave me interesting results. First, the coronal feild lines at maximum phases consist of many loops all over the globe so that the corona would look like a round ball as observed at maximum phases (see Fig. 3). The fields lines at minimum phases look like a dipole just like the real corona at minimum phases. At the same time, I noticed that the closed field lines at maximum phases and the open field lines at minimum phases can give us a unified interpretation of the solar cycle modulations of the galactic cosmic rays and the solar winds if the calculated coronal fields extend and fill out the entire heliosphere. Both cosmic rays and the solar wind are braked by the closed field lines at maximum phases but can stream in and out freely through open field lines at minimum phases. This gives us a sound basis of using ^{14}C data as deposited in the organic material on the Earth as a long-term indicator of the solar activity. At minimum phases, strong galactic cosmic rays produce many ^{14}C's in the Earth's atmosphere, while at maximum phases, the production is weakened. In fact, as Eddy analysed, the ^{14}C data revealed that the Sun underwent long-term modulations with time scales of hundreds and thousands years.

To discuss about the torsional oscillations with the same context of the linear

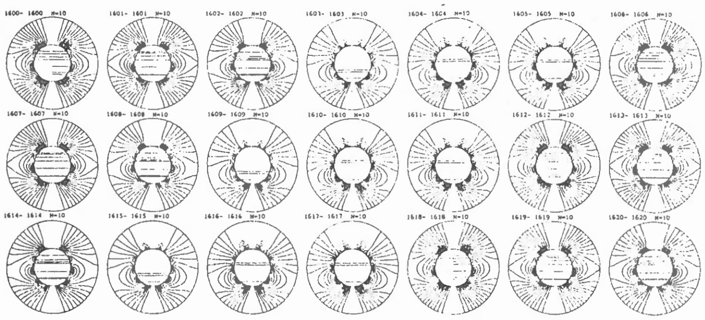

Fig. 3a. The dipole-like magnetic field lines of the corona at the minimum phase calculated from the surface fields of Fig. 2.

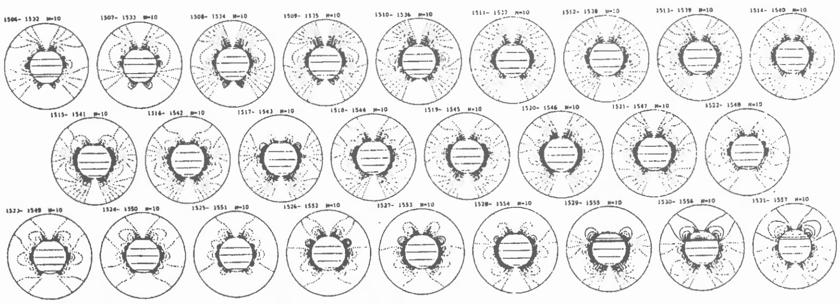

Fig. 3b. The round-shaped magnetic field lines of the corona with many loops during the maximum phase.

cases of the dynamo waves which are responsible for the characteristics of one cycle, I calculated the Lorentz force associated with the dynamo wave magnetic fields. The force consists of a nonoscillating part and an oscillating part, which I call the solar cycle Lorentz force waves. Only the nonoscillating part remains in the deep part of the convection zone. The oscillating part, which becomes conspicuous near the surface, has almost exactly the same pattern as that of the torsional oscillations (see Figs. 4 and 5). Rough estimate of excitation of the oscillations by the Lorentz force waves shows that the oscillations can in fact be driven by the force waves.

In this way, the characteristics of the solar cycle associated with one cycle can be understood in terms of the dynamo waves in the linear domain. The waves consists of a few mainly toroidal flux tubes that are successively created in the convection zone and are brought to the surface through formation of many active regions (see Fig. 5).

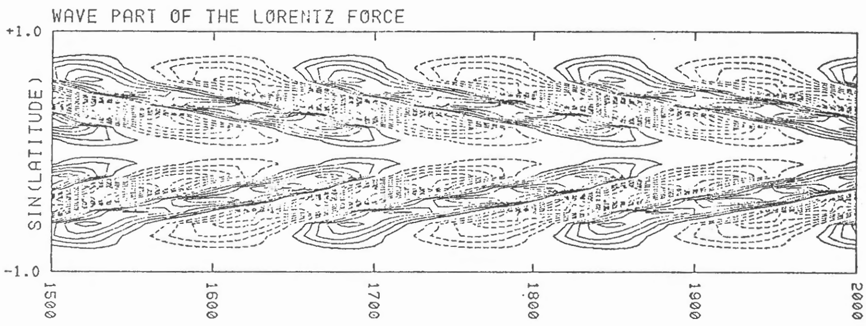

Fig. 4. The surface wave pattern of the Lorentz force of the dynamo waves of the solar cycle, simulating the pattern of the torsional oscillation. Ordinate is sin (latitude) and abscissa is time.

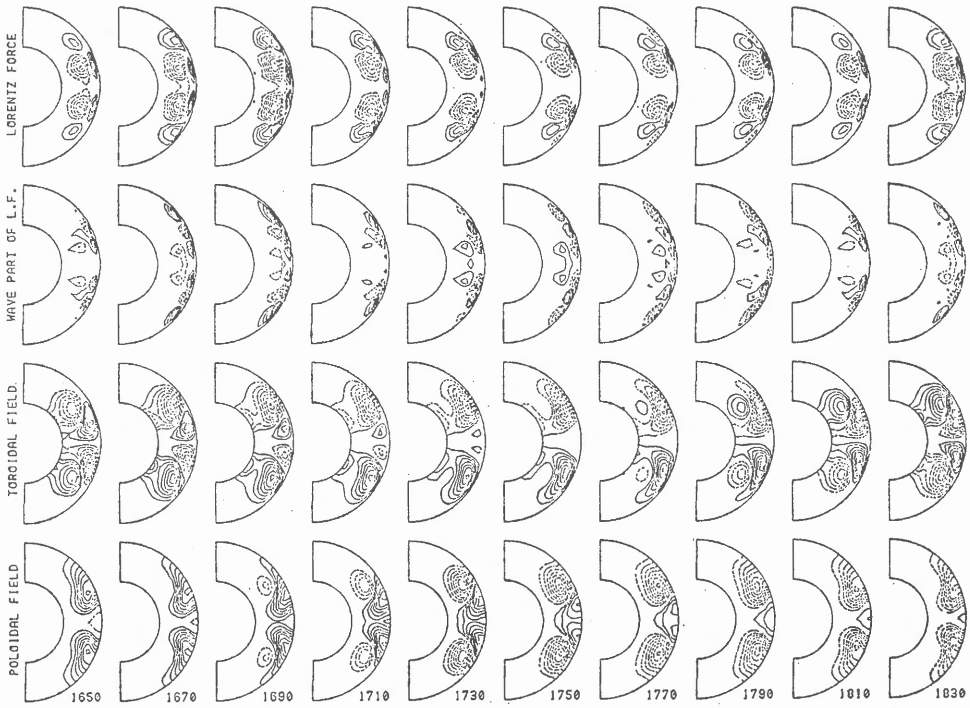

Fig. 5. Internal structures of the toroidal and poloidal general fields of the solar cycle, the longitudinal component of their Lorentz force, and its wave component. Numbers are time steps of the numerical integration of the dynamo equation.

2.4) Nonlinear Dynamo Problem

Next step is to study the long term modulations of the solar cycle. This is a nonlinear problem. There are two possible nonlinear mechanisms. One is through eruption of the magnetic fields which depends on the strength of the fields. But this possibility was eventually ruled out because of its inefficiency and because of the existence of the Maunder Minima. For, in the Maunder Minimum, there were few active regions, so that the nonlinear effects should have been small and the dynamo should have operated strongly, which leads to a self-contradiction, provided the dynamo kept operating at those times. Another possibility is that the dynamo strength depends on the magnetic fields. In order to estimate the effects, however, we need to solve the full system of equations with the Lorentz force of the zonal as well as nonzonal magnetic fields. Before doing this, I did some numerical experiments, assuming an experimental formula for the dependence of the dynamo number N_G and N_R as follows

$$N_{G,R} = N^0_{G,R} \exp \left(- \sum_{i=1}^{N} a_{Ni} \left| \Phi \, (t - t_{di}) \right|^{N_{fi}}_{MAX} \right) ,$$

The formula represents the suppressive feedback action of the field on the flows and eventually on the dynamo processes. Since it takes time for the magnetic fields to influence the flows, I assumed a time delay in the feedback. Some interesting results of this study are the following:

(i) A long-term modulation with 80-year or 55-year period naturally results in the solar cycle oscillations with the basic period of 11 years when the feedback is represented by a single delay time parameter. The theory predicted a hysteretic relation between periods and amplitudes.

(ii) Higher-order modulations with time scales of hundreds and thousands year periods can be superposed on the modulated oscillation when the feedback is represented by multiple delay time parameters. Especially, the Maunder-Minimum-like eras with few surface activity but with oscillating internal fields can be reproduced.

(iii) With no time delay, the solutions becomes either stationary oscillations or stationary steady states. This opens a way to study in parallel the solar-like stars with oscillating fields, as O. C. Wilson observed, and the stars with steady fields like the Earth and possibly early-type magnetic stars with convective cores.

(iv) With time delay, the steady field solutions display occasional polarity reversals as observed for the case of the geomagnetic reversals.

(v) The theoretical prediction of the hysteretic relation between the periods and the amplitudes led to the discovery of the 55-year grand cycle in the observed solar cycle. (see Figs. 6 and 7). This is a new basic property of the solar cycle which any model of the solar cycle must take into account.

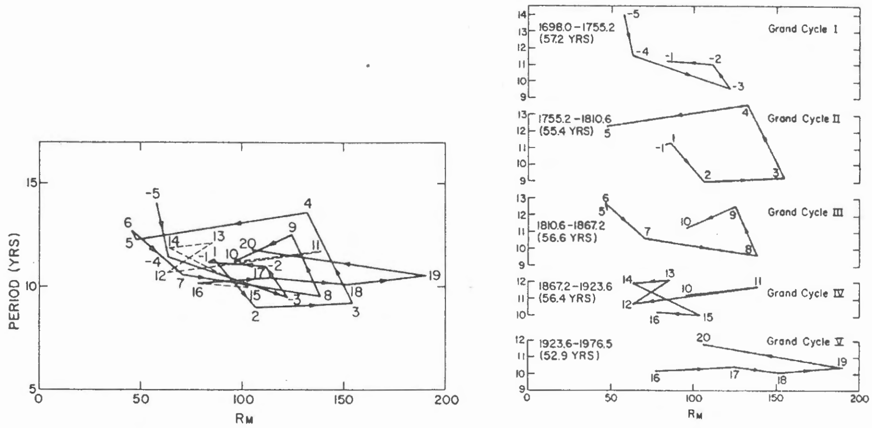

Fig. 6. Period-amplitude relation of the observed sunspot relative number curve. This shows that the solar cycle oscillation has a hysteresis and that the oscillation is undergoing the 55-year grand-cycle modulation.

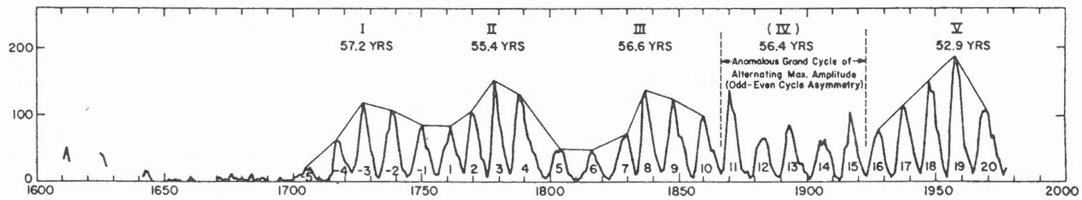

Fig. 7. The sunspot relative number curve divided into five 55-year grand cycles, each consisting of five 11-year cycles. Cycle 21 is the start of grand cycle VI.

In summary, we have a dynamo model to answer the five questions listed in the beginning of this talk. The magnetic fields of the Sun are created by the dynamo action of the differential rotation and the global convection. The oscillatory nature of the fields is a natural state of such dynamo processes. The global and local magnetic processes, associated with the dynamo mechanisms, which are operated by the differential rotation and the global convection, naturally cause the individual phenomena of the solar activity. The steady field-like geomagnetic field is a special state achieved by delicate nonlinear balance between the dynamo and the diffusion processes. This delicate balance is destroyed often and the field reverses its polarity in the oscillatory modes. The long-term modulations of the solar cycle with multiple-periods are also inherent and universal characteristics of the non-linear dynamo. With different Ω and R (i.e. with different convective structure and different dynamical processes), magnetic fields of other stars may also be studied within the same context. However, the theory as well as observation are still incomplete. Both are progressing now side by side.

References

Interested readers are referred to the general references cited in the following papers on which the present talk is based:

Publ. Astron. Soc. Japan **23**, 57-73(1971), **26**, 9-51(1974)

Solar Phys. **18**, 417-433(1971), **22**, 20-33(1972), **33**, 131-143(1973), **47**, 581-600(1976), **50**, 3-23(1976), **52**, 41-52(1977), **54**, 229-258(1977)

IAU Symp. **71**, 19-35(1976), **71**, 137-143(1976), **71**, 409-413(1976)

Ap. J. Suppl. No. 294, **29**, 467-494(1975)

Ap. J. **178**, 863-886(1972), **201**, 740-748(1975), **220**, 692-771(1978), **221**, 1088-1099 (1978), **226**, 706-719(1978), **227**, 1047-1058(1979), **235**, 625-649(1980), submitted (1980).