Transactions

of the

A.S.M.E.

MAY, 1940

Characteristics of Cloth Filters on Coal Dust–Air Mixtures

By A. R. MUMFORD,[1] A. A. MARKSON,[2] and T. RAVESE,[3] NEW YORK, N. Y.

This paper deals with the installation, operation, and maintenance of cloth filters for handling vented air from pulverized-coal mills. The experiences of the authors in solving problems in connection with the cloth-bag filter system, as used at the Kips Bay Station of the New York Steam Corporation, are the basis on which the subject is developed. Complete performance tests on a 40-ton-mill installation are cited.

TWO INSTALLATIONS of cloth filters have been successfully handling the vented air from two large pulverized-coal mills at the Kips Bay Station of the New York Steam Corporation for nearly three years.

In the application of such equipment to power plants, the magnitude and variation of the draft loss, when related to the frequency of the shaking cycle and to the capacity, assume greater importance than in smaller applications for air cleaning and, therefore, have been very carefully investigated.

Similarly the required maintenance is of greater importance because of the interdependence of the mill and other plant equipment and therefore this item has also been examined.

The presence of a fire hazard in the handling of combustible air mixtures adds to the complexity of the design and operation so the safeguards taken in the design and installation at this plant will be indicated.

DESCRIPTION OF PLANT

The Kips Bay steam plant was erected in 1926. The station was laid out for the storage system and this design was continued in the 1927 and 1930 extensions.

The millhouse contains seven pulverizing mills with a total capacity of 150 tons per hr. Raw coal entering the millhouse contains about $3^1/_2$ per cent moisture and of this 2 to 3 per cent is removed in the process of preparation. The coal is dried in external-grid driers, using exhaust steam for four mills while in three mills the coal is dried in the mill by air heated by high-pressure steam in air heaters. Air circulated through each mill system conveys the pulverized coal to the primary cyclone where separation of the coal and air takes place. The conveying air is vented to the atmosphere through a secondary cyclone and dust washer. The air-flow circuit for the 25-ton ball-type mill is given in Fig. 1.

One to two per cent of the coal milled may escape with the air vented from the primary cyclone and 0.25 to 0.5 per cent may escape from the secondary cyclone. For further reduction of the

discharge concentration, air washers have been generally used. The use of washers brought secondary problems which required continual attention. These are as follows:

1 Disposal of sludge water and the associated problem of maintaining drain lines free from clogging.
2 Corrosion of air washer and fouling of sprays.
3 Warm-air discharge, high in humidity, creates fog obstructing vision of roof operations.
4 Some spray water is discharged as a very fine mist, deteriorating building walls and creating a bad condition on the millhouse roof.
5 Loss of pulverized fuel as a result of the wet washing.

It is evident that, if the vent coal is to be reclaimed dry, a dry type of filter must be resorted to and further that such dry-filtering equipment must be nearly 100 per cent efficient under all load conditions or a potential nuisance will exist. One of the few types of filters on the market which fulfills the requirements is the cloth or bag filter system.

It was decided in 1936 to purchase two cloth-type filter collectors, each with four filtration compartments for the 25- and 40-ton per hr ball-type mills, eliminating from the vented-air circuit both the secondary cyclone and air washer.

The two filter units were installed under a single housing on the millhouse roof over the raw-coal bunker and adjacent to the south coal tower. The secondary cyclone and air washers on the ball-type mills were left intact after connecting the filter dust collectors so that, in the event of failure of the dry filters, which has not occurred to date, the vent system could be changed over to wet washing in a matter of hours. The dry-filter system air-flow circuit with the change-over provision for wet washing is shown in Fig. 2.

The dust filter compartments are in the same housing to sim-

DATA

1	RAW COAL BUNKER 4000 TON SECTION
2	RAW COAL FEEDERS
3	BALL TYPE MILL CAP 25 TONS PER HR
4	AIR HEATER FAN CAP 17500 CFM
5	STEAM AIR HEATER MAX AIR TEMP 410° F
6	PRIMARY CYCLONE COLLECTOR DIA 13 FT
7	TRANSPORT PUMP CAP 30 TONS PER HR
8	MAIN EXHAUSTER FAN CAP 17000 CFM
9	SECONDARY CYCLONE COLLECTOR - 9 FT. DIA
10	AIR WASHER
11	CLEAN AIR DISCHARGE
12	EXPLOSION VENTS

FIG. 1   COAL AND AIR FLOW CIRCUIT, 25-TON MILL WITH WET WASHING OF VENT

plify construction, although their operation is entirely separate. The path of flow of the coal-dust and air mixture through the collector is shown in Figs. 3 and 4. Flow of the mixture is up through the bottom of the collector, through a pipe located in the center of the filter then to the four compartments, two of which are located on one side of the inlet and two on the opposite side. At the inlet to each pair of filtering compartments is a gate-type damper which can be used to close off the flow to either compartment but never to both at the same time. Take-up springs are provided on the driving-shaft mechanism to prevent shearing of the damper shaft and also to insure a pressure fit of the damper against a $1/2$-in.-thick soft-rubber gasket which lines the door frame.

The dusty air enters the filtering compartment and makes a turn of 180 deg to the outside of the filtering surface. The filtering surface consists of a number of bags each shaped as a rectangular envelope open on one end. The bag or envelope is slipped over a galvanized-wire screen, the opposite surfaces of which are held in correct spacing of $1^3/_{16}$ in. by metal struts. The 25-ton mill has 552 such bags and the 40-ton mill 840 bags, six rows high. The path of air flow is through the cloth and out the open end of the envelope. The coal dust builds up a dust layer on the cloth surfaces, which is shaken off periodically, and the cleaned air passes through to an opening provided in the housing. The open end of each bag screen is sealed to the collector casing, dividing it into a dust side and clean side.

FIG. 2   COAL AND AIR-FLOW CIRCUIT, 25-TON MILL WITH CLOTH FILTERS
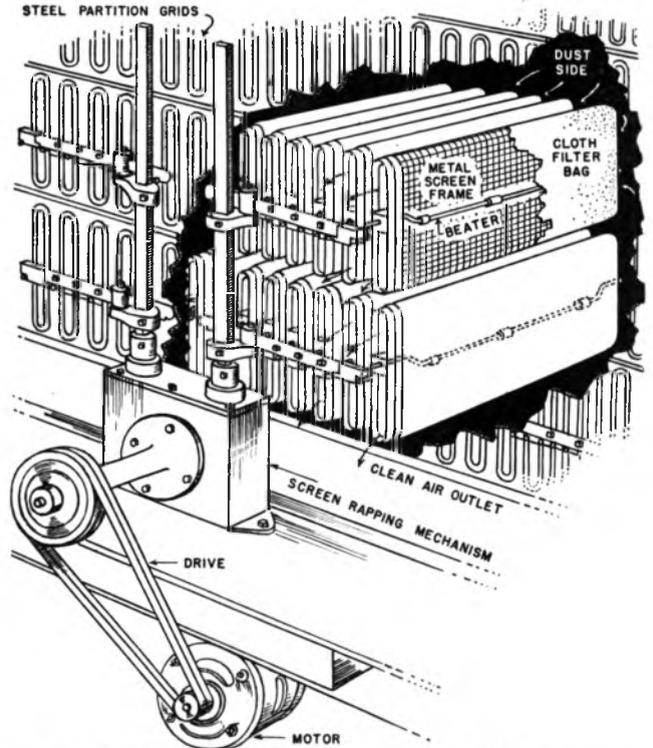
FIG. 4   ARRANGEMENT OF BAGS AND SHAKING MECHANISM
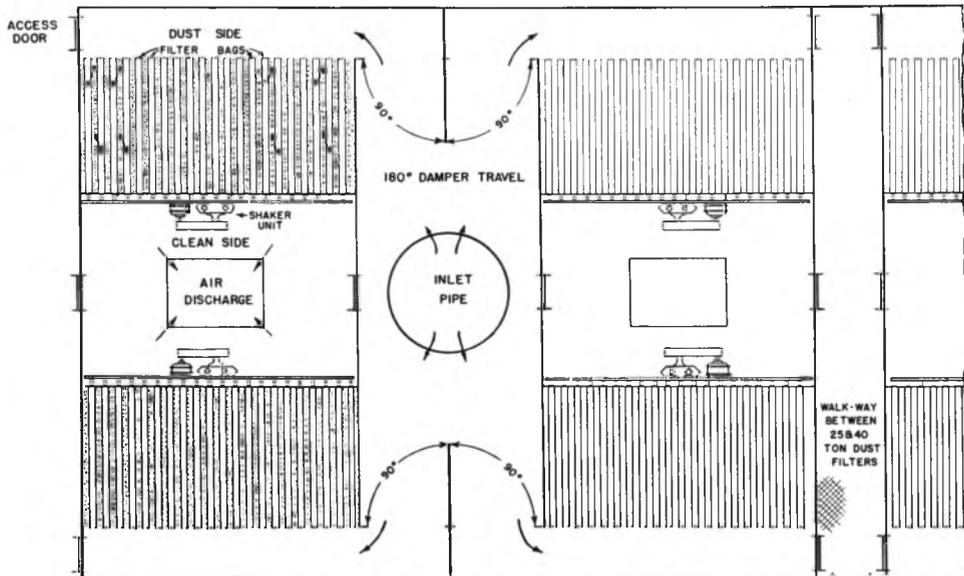
FIG. 3   AIR-FLOW DIAGRAM THROUGH DUST FILTER

The shaker consists essentially of a motor operating a rocker mechanism which transmits oscillating motion to beater rods located along the centers of the bag frames. Coal hoppers are located directly below the cloth surfaces to receive the loosened dust from the bags. A rotary valve, which is air-sealed by means of a rubber flap, rotates at slow speed and discharges the reclaimed coal to a return coal pipe which conveys it to the hopper above the transport pump for delivery with the main body of coal to the boiler fuel bins. Each compartment has its own shaker motor and shaking mechanism. The length of the shaking period and proper sequence is governed by the cam switches on the time-control panel.

LABORATORY TESTS

The successful installation of the collector depended on several factors, some of which had to be determined in the laboratory because the information was unavailable. The known factors were:

1 Vented-air volume.
2 Vented-air temperature and relative humidity.
3 Vented-coal quantity and screen sizing.
4 Maximum pressure loss allowable in operation.

The factors which had to be investigated because little or no information was available were:

1 Pressure-drop characteristic of the filter cloth.
   (a) The effect of flameproofing on the pressure drop.
   (b) The effect of coal-dust loading on the pressure drop.
2 Safety considerations.
   (a) Precautions necessary to insure safe operation of the cloth filters.

The known quantities as previously established are given in Table 1.

TABLE 1   KNOWN FACTORS OF COLLECTOR OPERATION

	25-Ton mill	40-Ton mill
Vented-air volume, cfm, max	15000	27500
Vented-air temperature, F	130	130
Vented-air, relative humidity, per cent	60	60
Vented coal, lb per hr, max	800	1600
Vented-coal sizing	100 per cent through 325-mesh Tyler screen	
Available pressure for cloth filters, in. water, max	6	6

Tests were made on a number of filter cloths submitted by manufacturers. The cloths were tested for air-flow resistance in the clean condition and also after loading with coal dust. The fabrics tested were made from cleaned cotton having a sateen weave with a count 96 × 64 threads per in. Two weights of fabric tested are reported; one with a weight of 1.05 yd per lb and the other 1.30 yd per lb; both fabrics 54 in. wide. The cloth samples were clamped tight without stretching to one end of a cylindrical chamber approximately 10 in. diam and 3 ft long. Metered clean air was blown in on the other end. Cloth resistance was measured with a U gage connected to the chamber approximately 2 in. from the cloth. The effect of several flameproofing solutions on the pressure drop was also established. The designations of cloth tests reported are given in Table 2.

TABLE 2   DESIGNATIONS OF CLOTHS TESTED

Curve symbol designation	Cloth thread count	Cloth width, in.	Cloth weight, yd per lb	Flameproofing
A	96 × 64	54	1.05	Untreated
A_1	96 × 64	54	1.05	Mixture of ammonium-phosphate and boric-acid solution
B	96 × 64	54	1.30	Untreated
B_1	96 × 64	54	1.30	Sodium-borophosphate solution
B_2	96 × 64	54	1.30	Sodium-ammonium-borophosphate solution
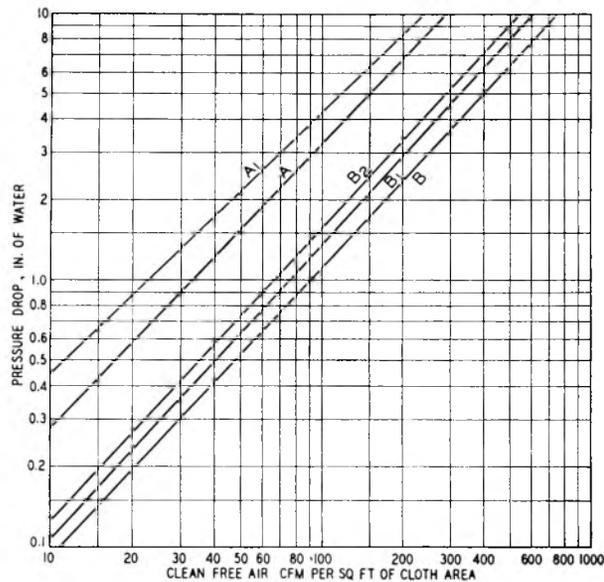
FIG. 5   PRESSURE DROP THROUGH CLOTH WITH CLEAN AIR
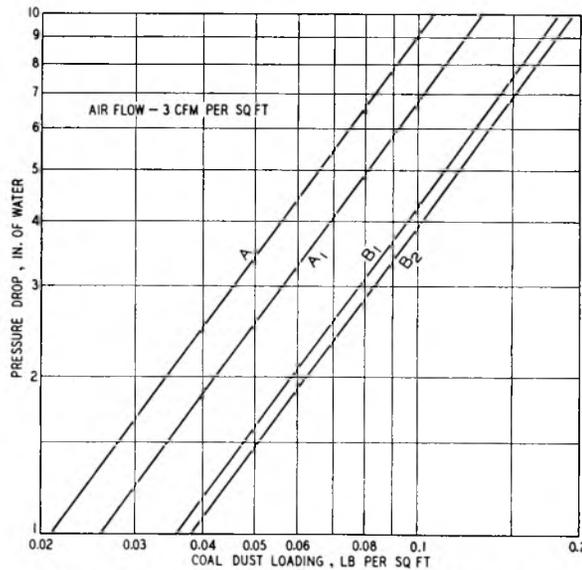
FIG. 6   PRESSURE DROP THROUGH CLOTH LOADED WITH COAL DUST

The results for clean free air are given in the curve in Fig. 5 and indicate that the flow-resistance curves are parallel, with the exception of curve A_1 which may be accounted for by the fact that cloth A_1 may have had an initial stretching.

The difference in the weight of the cloths A and B is reflected in their relative resistance to air flow. Cloth B (the lighter cloth) will pass approximately 2.7 times as much air as cloth A for the same pressure drop. Flameproofing increases the pressure drop of cloth A by about 24 per cent and that of cloth B by about 18 to 25 per cent. The effect of flameproofing on the flow resistance will depend largely on the composition, strength, and also the method of application to the cloth.

Continuing the foregoing test; some fine coal (screened through a 200-mesh Tyler screen) was blown against the cloth and pressure-drop data collected. These data were taken for a number of coal loading points so that it was possible to study the rate at which the pressure drop increased. Powdered coal was intro-

duced into the cylindrical chamber mentioned with an air injector, making it possible to secure a uniform coating of coal on the cloth, the amount of which was determined by weighing the cloth after each run.

The results of these tests are indicated in Fig. 6, which shows the effect of coal loading on the pressure drop for an air flow of 3 cfm free air per sq ft. The curves are plotted for cloths A, A_1, B_1, and B_2 and have the same letter designation given in Fig. 5. If, as an example, for cloth A a point at 0.05 lb per sq ft coal-dust loading is taken the pressure drop is 3.4 in. Comparing with Fig. 5, a pressure loss of 3.4 in. will pass 102 cfm free air per sq ft or 34 times as much air. It will be noted that, in line with the tests of Fig. 5, cloths B_1 and B_2 will allow more coal accumulation on the cloth for the same pressure-drop and flow conditions than cloths A and A_1. Thus for a 3-in. pressure drop, cloth A_1 will allow a coal loading of 0.056 lb per sq ft, whereas cloths B_1 and B_2 will allow 0.078 and 0.083 lb per sq ft, respectively. This, how-
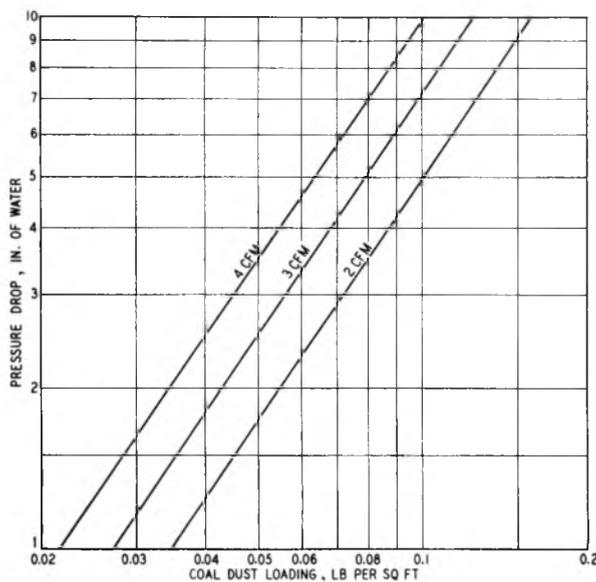
FIG. 7   PRESSURE DROP THROUGH CLOTH A_1 VERSUS COAL-DUST LOADING

ever, does not mean that the normal operating pressure of the cloth filters will be less with cloth B_1 or B_2 than with cloth A_1, but rather that cloths B_1 and B_2 will retain more coal on their surfaces than cloth A_1 and that the residual loading, after the cleaning period of cloths B_1 and B_2, is more than with cloth A_1. The normal operating pressure is approximately the same with either cloth as borne out by actual operating experience with both cloths.

Observations made during these tests indicate that coal dust passes through the cloth when it is clean until the pores become filled with dust particles to a point when no more coal dust passes through. The dust retained on the cloth provides a mat filtering medium. This effect is more noticeable with cloths B_1 and B_2

TABLE 3   CLOTH-FILTER SURFACE AS DESIGNED

|	25-Ton mill	40-Ton mill
Coal milled, lb per hr	50000	100000
Coal to cloth filters, lb per hr	800	1600
Vented air, cfm	15000	27500
Free air cfm per sq ft cloth area (100 per cent area assumed)	2.1	2.6
Free air cfm per sq ft cloth area (75 per cent area)	2.8	3.5
Total cloth area, sq ft	7040	10710
Reduced area, 75 per cent of total, sq ft	5280	8030
Coal-dust loading, lb per M sq ft to give 4 in. normal operating pressure at reduced area	71	62
Total coal on cloth filters, lb	500	665
Shaking interval, min	37	25

than with cloth A_1. No tests were made with cloth B (untreated cloth) because it had been decided to use flameproofed cloth in the installation.

Cloth A_1 was initially selected for the installation because of its superior mechanical properties. It was believed that the heavier cloth would withstand the effects of temperature and beater rods better than the lighter cloths B_1 or B_2. Flow curves were constructed for cloth A_1, as in Fig. 7.

The design for cloth-filter surface for both mills is contained in Table 3.

From the figures given in Table 3, it is evident that the cleaning-period cycle should be less than 37 min for the 25-ton mill and less than 25 min for the 40-ton mill. It was recognized that the calculated figures are only approximations because it had been assumed that the filter cloths were thoroughly clean at the beginning of each cycle. The limited value of the tests, because of the number of uncontrolled variables, such as temperature, humidity, dust and sizing, did not justify more accurate determination of the cleaning cycle. To simplify the control for these operations, it was decided to operate both mills with the same initial cleaning cycle, a compromise between the two calculated cleaning cycles.

Each filtering unit has four compartments and is so constructed that one compartment at a time is removed from service for cleaning. It is, of course, important that the cleaning operation occur only while the compartment is dampered off. The setting adopted is given in Table 4.

TABLE 4   CYCLE SETTING ADOPTED

	25-Ton mill	40-Ton mill
Number of complete cycles per hr	2	2
Time required for 1 cycle, min	30	30
Total time, unit operating with 100 per cent filtering capacity, min	10	10
Total time per cycle when unit is operating with 75 per cent filtering capacity, min	20	20
Actual shaking time per compartment per cycle, min	3	3
Actual shaking time per compartment per hr, min	6	6
Actual shaking time all compartments per cycle, min	12	12
Actual shaking time all compartments per hr, min	24	24

NOTE: Time control for the cycle was obtained from a synchronous motor through reduction gears driving a shaft onto which were fastened bakelite cams opening and closing contact switches in proper sequence for damper and shaker motor operation. Provision was also made for manual operation of the control limit switches.

The validity of such testing with respect to the resistance of different cloths has a good degree of plausibility, since it is possible to test the cloths under practically identical Reynolds-number conditions as exist in the full-scale apparatus. These tests indicate that the effective filter resistance is principally determined by the filter mat which is formed by the cloth and the so-called residual dust loading. To a considerable degree, independence of filter resistance to the textile specification of the cloth exists, a fact of interest because cloths may be selected principally for their mechanical qualities.

If it should be found that such tests furnish an approximate basis for the design, they would assume considerable importance. The authors' results indicate that this is only roughly true. Since such correlation was not of primary interest at the time of carrying out the work, necessity for further and more extensive data is indicated, especially on different materials.

The laboratory tests indicate two apparent flow-resistance rules. First, at constant loading of the cloth the draft loss varies directly as the flow. This is reasonable enough as the flow through the filter itself must be viscous. Second, at constant air flow, the draft loss varies as the 1.5 power of the loading, a statement by no means so obvious and which must be confined to the limits of the experiments, because it involves several variables

connected with dust sizings, atmospheric conditions, and others not under controlled variation in the experiment.

FIELD TESTS

The installation of the 25-ton-mill filter was completed in the spring of 1937 and placed in service immediately thereafter. For the first several days, operation of the collector was as expected. Initially some coal dust came through the cloth onto the clean side until a dust mat was formed on the cloth after which the air coming through the cloth was perfectly clean. Subsequent operation of the collector showed that the air resistance kept building up and no amount of shaking would bring it back to normal while the mill was in service, although the shaking was effective with the mill shut down. After numerous tests, it was decided to change the timing cycle. New cams were ordered, cut, and installed, which permitted more flexibility in the time-control variation by resetting the cams on the drive shaft and changing the motor reduction gears.

During the shaking cycle, the sequence of the cleaning operation is as follows:

1 Close off compartment to filtering by damper.
2 Time prior to shaking called "rest period."
3 Shaking period.
4 Time after shaking called "settling period."

Changes were made in the number of cleaning cycles per hour and the rest, shaking, and settling periods.

Some of the results of the tests have been plotted in curve form
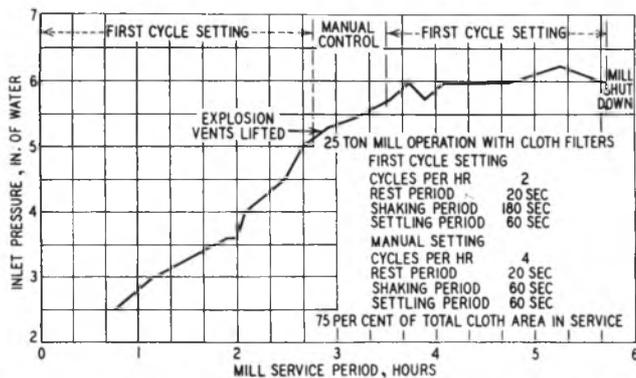
FIG. 8 VARIATION OF BACK PRESSURE WITH TIME; FIRST-CYCLE SETTING

and indicate the rate of pressure build-up. Fig. 8 is a typical curve of the initial cycle for 75 per cent filtering area. During part of this cycle 100 per cent filtering area was used. However, so that the test results of the original cycle may be compared with subsequent data on cycles which employ 75 per cent filtering area all of the time, the results were plotted for 75 per cent filtering area. In a period of 2 hr 5 min pressure built up to 4 in., the normal design pressure, and increased to 5¼ in. in an additional 47 min, resulting in a lifting of the explosion vents. After 2 hr 45 min of service the time control was changed to manual with 4 cycles per hr and 60 sec shake, but pressure continued to build up until the vents were continually open. The mill was shut down after 5 hr 45 min of service because of excessive back pressure.

Figs. 9, 10, and 11 are plots of results of the tests in which the cleaning cycles per hour, rest, shaking, and settling periods were changed as indicated on the curve sheets. All tests indicate the filter back pressure rises to 4 in. from 2 to 3¼ hr and, with the exception of one test (Fig. 10) in which 4 in. is reached in about 1½ hr, continued to increase until the vents lifted at 5¼

in. and the mill was shut down. In each run the milled coal was 21 tons per hr, air flow approximated 14,000 cfm, dry-bulb temperature 125 F, and relative humidity 40 per cent. No effect of the cycle frequency on the filter back pressure is apparent except that the results with the 3-cycle frequency are worse than with
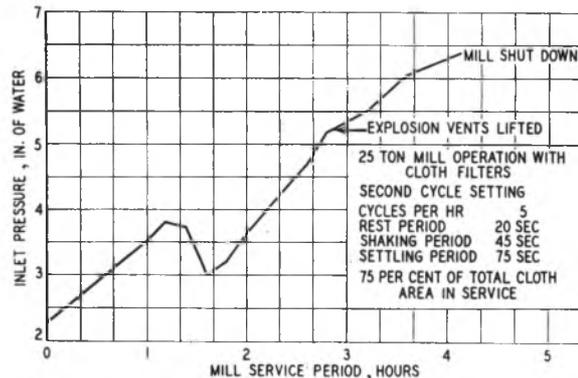
FIG. 9 VARIATION OF BACK PRESSURE WITH TIME; SECOND-CYCLE SETTING
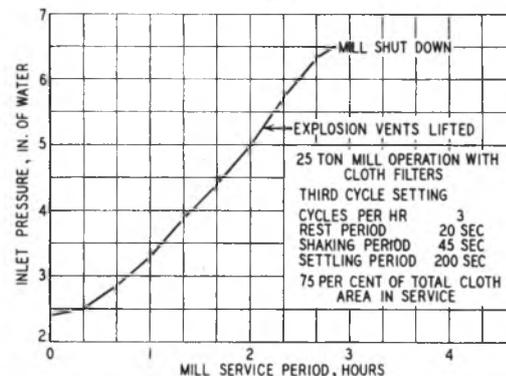
FIG. 10 VARIATION OF BACK PRESSURE WITH TIME; THIRD-CYCLE SETTING
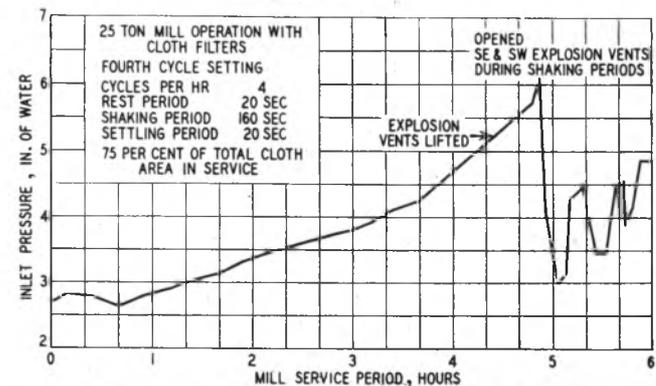
FIG. 11 VARIATION OF BACK PRESSURE WITH TIME; FOURTH-CYCLE SETTING

the 4 or 5 cycle. The duration of the shaking period had no effect on the pressure build-up.

Continued checking of the shaker mechanism and damper-operating mechanism indicated them to be in good working condition. When the damper closed off a compartment to air flow, preparatory to shaking, there was a gradual drop in pressure by virtue of the fact that flow occurs until the pressures on both sides of the filter cloth are in balance. If the time allotted for pressure balance is not sufficient or is offset by air leakage around

the damper seal, then shaking of the cloth no matter what the duration will be ineffective. A checkup of the compartment pressure, when dampered off, indicated a positive pressure of approximately 0.02 to 0.04 in. of water.

During the test run plotted in Fig. 11, two of the vent doors were lifted accidentally while the compartments were dampered off, thus allowing quick equalization of clean- and dirty-side pressures and also providing a sufficiently large area for escape of any inleakage without affecting the pressure balance. The effects were readily apparent. The filter resistance dropped from 6.10 to 4.10 in., after one of the compartment filters had been cleaned, and dropped to 3 in. after the cleaning of the second compartment filters. The pressure increased to 4.4 in. as the first then the second compartment on which the vent doors had been lifted were removed from service. The pressure dropped from 4.4 to 3.45 in. when both filtering compartments were returned to service.

DESIGN OF RESIDUAL-PRESSURE RELEASE SYSTEM

When the pressure filter system, as previously indicated, went into operation the resistance built to very high values. Increase
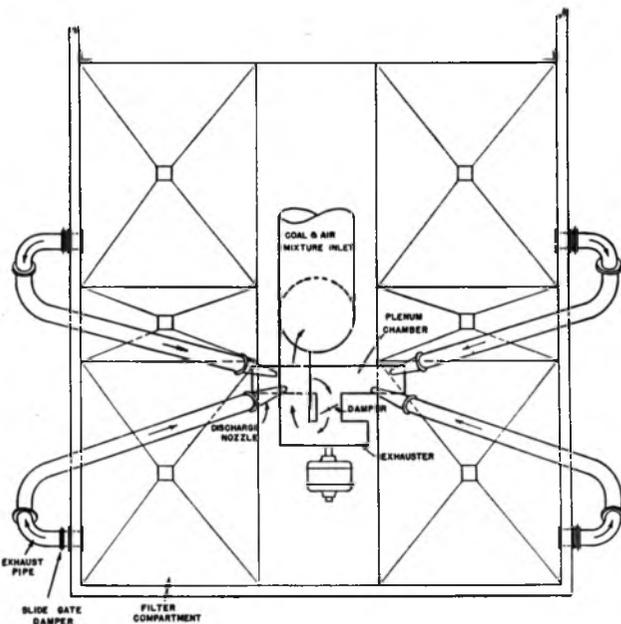
FIG. 12 EXHAUSTER-FAN ARRANGEMENT FOR OPERATION OF DUST FILTER

of the number of shaking periods and of the time of shaking was ineffective in eliminating pressure build-up. One theory advanced was that the residual air pressure in the isolated compartment held the coal against the filter in spite of efforts to shake it free. Tests were run to determine the amount of evacuation necessary to allow the free discharge of the dust from the cloth. It was found by measurement of air flow that 275 cfm were required for the filters of 410 cu ft free volume, and 400 cfm for the filters of 590 cu ft free volume. A margin for damper leakage was also allowed.

After consideration this evacuation was carried out in the manner shown in Fig. 12. A single fan was connected to all compartments and continuously operated. Its capacity requirements were based upon the following criteria:

The operating suction at the fan inlet is selected so that variations in flow from operating and nonoperating compartments owing to differences in pressure will be swamped out to a desired degree. This necessitates the introduction of accurate resistances

in the individual lines, which may be tapered nozzles, properly designed and discharging in the inlet plenum. The over-all system resistances including the nozzles should be calculated because high line velocities should be used. The total flow into the fan under operating conditions is readily figured and leads to the selection of a fan of proper P-V characteristics. Low velocities and flat spots must be avoided as shown in Fig. 12.

The power requirements of this fan are a charge against operation amounting to 7 hp for the 7040-sq ft filter and 12 hp for the 10,710-sq ft filter.

The necessity for some such installation would exist whether the filter were under pressure or suction. However, for a suction filter, a simple mechanical atmospheric break on the clean side might be provided for each compartment.

The reasoning indicating the necessity for evacuation is expressed mathematically in the following expression which offers a means for evaluating the time which must elapse before the particles may be shaken off in the face of a slight residual pressure.

If Q_t, $\overline{V}$, P_t are, respectively, the weight of air in the filter, the volume of the filter, and the absolute filter pressure at any time t

$$Q_t = k_1 \, \overline{V} \, P_t$$

and the time derivative is

$$\frac{dQ_t}{dt} = k_1 \, \overline{V} \, \frac{dP_t}{dt}$$

the rate of outflow from the compartment is

$$-\frac{dQ_t}{dt} = k_2 \, \frac{P_t - P_a}{X} = k_3 \, (P_t - P_a)$$

in which $\dfrac{P_t - P_a}{X}$ is the pressure gradient across the filter cloth, P_a being atmospheric pressure.

$$k_3 \, (P_t - P_a) = -k_1 \, \overline{V} \, \frac{dP_t}{dt}$$

from which $- T = K \log (P_t - P_a) + C$

or $T \Big|_{t_1}^{t_2} = K \log \left(\dfrac{P_1 - P_a}{P_2 - P_a} \right) = K_4 \log \left(\dfrac{V_1}{V_2} \right)$

where V_1 and V_2 are the velocities through the filter mat.
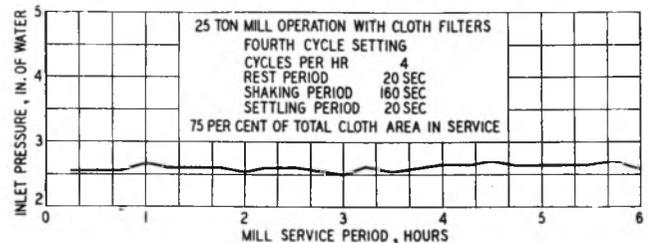
FIG. 13 VARIATION OF BACK PRESSURE WITH TIME; RESIDUAL-PRESSURE RELEASE SYSTEM

If the time for decrease of the pressure to $1/2$ value is observed, the time for the volume rate discharge to change from V_1 to V_2 is given simply enough by the foregoing ratio, assuming that the inlet damper is tight. Theoretically, at least, if the terminal velocity of the particles were known, the time the filter was to be out of service before the pressure decreased to the point at which the particles would fall off the cloth would be determinate.

A consideration of the terminal velocity of fine particles, as discussed by Croft,[4] leads to the general conclusion that the air velocities for the particles encountered here must be less than about 3.5 in. per sec and may have to be as low as 0.001 in. per sec before shaking can be effective. This consideration indicates the necessity for residual-pressure release.

The effect of the residual-pressure release mechanism on the back pressure is clearly shown in Fig. 13 which should be compared with Figs. 8, 9, 10, and 11.

INSPECTION AND MAINTENANCE

Of particular interest in the operation of the dry dust filters is the number of man-hours involved in inspection and maintenance. Inspections while the equipment is in service cover the following:

1 Lubrication of bearing surfaces of coal-conveyer valves, evacuating fans, and motors.
2 Checking housing for leaks from dust side of collector.
3 Observation of clean-air discharge for evidence of torn or defective bags.
4 Checking reclaimed-coal lines for free movement of coal.
5 Checking operation of main exhauster damper control.
6 Checking pressure balance of cloth filter with draft gage.

Items 1 to 4 are covered by the regular millhouse operating crew and are made a routine matter along with other duties at similar locations about every 2 hr. The inspection is primarily a precautionary measure. Items 5 and 6 are looked after by test men in the technical department of the company. A monthly inspection of the damper control is all that is generally necessary and requires about 3 or 4 hr per month. Item 6 is included for the purpose of checking the auxiliary-fan operation and is made about every 3 or 4 months in about $1/2$ hr.

During the shutdown operation it is advisable to check the following and do general internal cleaning:

1 Shaker mechanism.
2 Inspect damper gaskets.
3 Inspect auxiliary-fan piping.

It is good practice to make a general internal inspection of the collector weekly.

Maintenance data on the dust filters are not as complete as they should be at this time because included in the maintenance figures are design changes. These changes in design will undoubtedly lower maintenance charges and decrease outage periods in the future. Figures subsequently cited are for the total operating hours given for each filter.

The largest single item of maintenance in the dust filters is the filter bag. Bag life is approximately 2 yr and the bags, with flameproofing applied, cost approximately $1 each. To remove and replace bags requires about 180 man-hr for the 25-ton mill and 275 man-hr for the 40-ton mill, assuming all bags are replaced. Defective bags are located from the clean side of the collector and are replaced from the dust side of the collector. Since there are 552 bags in the 25-ton-mill installation and 840 bags in the 40-ton-mill installation, this replacement is therefore an appreciable item and worthy of investigation.

At the present time, tests are under way to confirm a proposed decrease in shaking frequency and time. One other change has been made which should improve the life of the bags. Formerly the flow of clean air was downward but now the flow is upward through the top, in order to prevent clean air with high vapor

content passing from one collector to the other when one of the collectors is out of service.

The next largest item of maintenance is general cleaning and gasket repair. Approximately 400 man-hr have been spent in this work but it is reasonably certain that this item will virtually disappear, because 90 per cent of the time devoted to cleaning would not have been necessary if the collector were dust-tight initially. For example, the access doors and coal-hopper casing leaked dust continually while the collectors were in service and required constant cleaning and patching, which otherwise would not have been necessary. Dust-tight access doors were made to replace the initial doors and the hopper casing was welded to make it dust-tight. Gasket maintenance is a relatively small item and has approximated 50 man-hr, which time is spent about once a year renewing gaskets on explosion vents, access doors, and damper frames.

Reclaimed coal drive systems have had approximately 330 man-hr maintenance, a considerable portion of which can be considered as due to the dusty atmosphere, resulting from leakage which in turn caused bearing maintenance and maintenance of the chain drive. Rotary valves which feed the coal into the screw conveyer and return coal pipe have had no maintenance. Some trouble had been experienced in the early days with flooding of the conveyer screw, causing plugging of the return coal pipe at the point where it connects to the screw, but this has been rectified by increasing the angle of inclination of the outlet coal pipe. At the present time the coal drive system runs continuously, but it is planned to try out intermittent operation of the drive system which, if successful, will decrease maintenance.

The shaker mechanism has needed 205 man-hr of maintenance which has been due largely to wear of the roll which fits in the cam groove of the rocker mechanism. The roll was worn egg-shaped because it did not rotate with the cam but slid in the cam groove. Some of the rolls were worn down to the pin in a period of 6 to 8 months but, when a grease-seal ball bearing was substituted as a cam follower in place of the hardened steel roll, very slight uniform wear resulted in a period of 18 months. There has been some loosening up and falling out of setscrews fastened to the rocker shafts which have been replaced with a self-locking type of setscrew.

Approximately 75 man-hr have been charged to resetting limit switches. This is not excessive, since it is good practice to check the limit switches on the gate-damper motors about once every 6 months. The purpose of the check is to insure a good seal of the damper against the door frame at all times so that the pressure balance of the cloth will be proper for effective shaking.

The 25-ton mill has been in service 6197 hr and the 40-ton mill 5759 hr since the installation of the dry filters. During this time it is estimated that 4400 tons of coal have been reclaimed by the dust filters. The operation of the filters has been normal and no major incidents in their operation have occurred.

SAFETY MEASURES

It is recognized that coal milling is a process which must be carefully handled to prevent fires during the operation. Protective equipment and safe procedures had been employed in the millhouse in line with best practice and this policy was extended to the dry filters. Numerous tests on filter cloth, both treated and untreated with flameproofing, indicated the desirability of treating the cloth, although it was recognized that such treatment would increase the flow resistance of the cloth slightly and perhaps affect its life. The effect of the flameproofing was found to be that of retarding the burning of the cloth and localizing it, although the time required for ignition of the treated and untreated cloth seemed the same.

Long horizontal runs of pipe are objectionable as are low pipe

---

[4] "The Calculation of the Dispersion of Flue Dust and Cinders From Chimneys," by H. O. Croft, Trans. A.S.M.E., vol. 57, 1935, paper FSP-57-1, pp. 5–10.

velocities. Accordingly in the layout of the pipe lines leading to the collector, the velocities are maintained at above 3000 fpm and the pipe runs are all practically vertical with the exception of the tie-in connection to the collector on the 40-ton mill, which is approximately 25 ft long. Exposed pipe above the roof is well insulated and provided with waterproof covering. In addition the horizontal run is provided with an inspection and access door. Dry-coal return lines are vertical with minor exceptions, where 45-deg bends were necessary. These were provided with cleanout plugs.

It has been mentioned that the dust filters were installed on the millhouse roof. The collector, therefore, had to be well insulated to prevent condensation. A 4-in. layer of rock wool backed up with $1/_4$-in. transite sheet was provided, the collector itself being lined with No. 14 gage sheet metal. The temperature drop through the collector is therefore small, in the order of 10 F, and the exit-air temperature is above the dewpoint.

Each compartment is provided with an explosion vent in addition to those already existing on the mill proper. Each filtering compartment has an access door.

Published data on the subject of inflammability of coal dust-air mixtures are meager with reference to the problem at hand. Two general statements may be made:

1 The dust must be present in a cloud of inflammable density and composition.

2 There must be a source of ignition, such as freely burning coal or an electric spark.

These two basic conditions may be discussed in terms of the factors which determine them. With regard to any distinction between inflammability and explosiveness, investigators of the Bureau of Mines[5] find, as a practical matter, that it is impossible to distinguish between them and that it is inadvisable to attempt such distinction.

When the coal-air mixture at Kips Bay Station leaves the primary cyclone, the velocities and composition are as given in Table 5.

TABLE 5 VELOCITIES AND COMPOSITION OF COAL-AIR MIXTURE, LEAVING PRIMARY CYCLONE

	25-Ton mill	40-Ton mill
Grains of dust per cu ft of air at 130 F........	6.2	6.8
Velocity in pipe, ft per min.................	3300	3400
Dust fineness............................	(100 per cent through 325-mesh Tyler screen)	

The question of mill drying by flue gases was studied at length and abandoned as economically impractical in this instance.

The regulation of air flow from the mill to the filters is accomplished by a high-quality automatic control which throttles the mill fan discharge to maintain a constant suction at the mill cyclone outlet. This function, the maintenance of constant mill air flow, has performed admirably with concurrent advantages in mill operation. Variations in filter pressure are not felt in the mill system.

On the receiving hoppers of the filter, mechanical rappers were installed and have worked very well. The rotary valves for discharging the hoppers are mechanically driven and isolate the hoppers at all times from the receiving screw conveyers.

The receiving screw conveyers were the scene of several fires in the early operation due to spontaneous combustion of packed coal in the barrels. This condition was eliminated by minor changes. Such precautions are also directed against fire due to spontaneous combustion.

---

[5] "Coal-Dust Explosibility Factors Indicated by Experimental Mine Investigations, 1911–1929," by G. S. Rice and H. P. Greenwald, U. S. Bureau of Mines, Washington, D. C., Technical Paper No. 464, 1929.

It is also necessary to insure against the hazards of static electric discharges. This insurance is largely inherent in the fact that this filter is a mass of metal mesh supporting the cloth, giving almost continuous grounding with a minimum free volume for the occurrence of intercloud discharges.

There is some evidence that high humidities reduce the inflammable character of the mixture, although publications of the Bureau of Mines declaim the futility of relying on this. For instance, Rice and Greenwald[6] unequivocally reject the influence of humidity in practical work. However, high humidities may reduce static-electricity hazard considerably.

Appendix

Three rather complete performance tests on the 40-ton-mill cloth-filter installation are given in Table 6. Of particular interest is the power consumption of the cloth filters and the fineness of the reclaimed coal. Power consumption of the cloth-filter installation is less than 0.2 kw per ton of milled coal, which is approximately 1.5 per cent of the total. Fineness of the reclaimed coal indicates that practically 100 per cent passes through a 325-mesh Tyler screen and also that the density of the reclaimed coal is approximately one half that of the milled coal. It is of interest
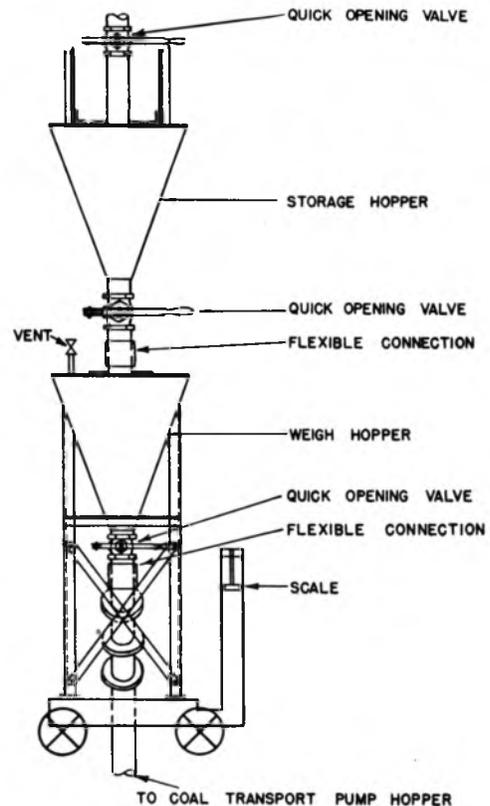
FIG. 14 RECLAIMED-COAL WEIGHING DEVICE

to note that the efficiency of the primary cyclones is better than 99 per cent.

The tests were made with 4 cleaning cycles per hr on the cloth filters. The duration of each shaking period was approximately 20 sec per compartment for each cleaning cycle.

The reclaimed-coal weighing device used in these tests is indicated in Fig. 14. The particular advantage of the installation

---

[6] See reference (5); p. 8.

TABLE 6 PERFORMANCE TESTS OF 40-TON MILL WITH CLOTH FILTERS

	June 17, 1938	June 20, 1938	June 21, 1938
Date			
Test no.	2	3	4
Duration, hr.	4.53	4.02	5.0
Total coal milled, tons	192.2	150.3	188.1
Coal milled per hr, weighed, tons	42.35	37.35	37.62
Power Consumption:			
Mill including heater fan, kwhr	377.0	371.8	381.8
Feeders and magnetic separators, kwhr	1.8	1.8	2.0
Main exhauster fan, kwhr	112.8	111.0	112.0
Transport pump, kwhr	9.8	10.8	9.6
Cloth-filter motors, inc. conveyer, kwhr	7.0	7.0	7.0
Total, kwhr	508.4	502.4	512.4
Mill including heater fan, kw per ton	8.90	9.95	10.13
Feeders and magnetic separators, kw per ton	0.04	0.05	0.05
Main exhauster fan, kw per ton	2.66	2.98	2.98
Transport pump, kw per ton	0.23	0.29	0.26
Cloth-filter motors, including conveyer, kw per ton	0.17	0.19	0.19
Total, kw per ton	12.00	13.46	13.61
Air-Heater Data:			
Fan-inlet temp, dry bulb, F	80	87	88
Relative humidity, per cent	50	38	44
Discharge pressure, in. water	8.9	9.0	9.0
Fan speed, rpm	1765	1760	1767
Heater-inlet pressure, in. water	3.7	3.6	3.6
Heater-outlet pressure, in. water	2.7	2.8	2.7
Heater pressure drop, in. water	1.0	0.8	0.9
Heater outlet temp, F	236	234	227
Hot air to mill, cfm	20700	20700	21100
Hot air to mill, lb per hr	69300	69200	71600
Steam to heater, lb per hr	2530	2380	2270
Btu available by steam in heater, M Btu per hr	2430	2381	2200
Heat-up by fan, M Btu per hr	47	47	49
Total Btu input, M Btu per hr	2477	2428	2249
Btu absorbed by air, M Btu per hr	2590	2440	2390
Radiation and unaccounted for, M Btu per hr	−113	−12	−151
Radiation and unaccounted for expressed as per cent	−4.7	−0.5	−6.7
Mill Data:			
Exhaust temperature, F	115	118	116
Pressure air entering mill, in. water	1.7	1.8	1.5
Pressure before classifier, in. water	−6.4	−6.6	−6.6
Pressure exhaust, in. water	−9.4	−9.6	−9.7
Air leaving mill, lb per hr	70800	72300	70800
Air inleakage, lb per hr	1500	3100	−800
Air inleakage expressed as, per cent	2.2	4.5	−1.1
Pressure drop across classifier, in. water	3.0	3.0	3.1
Pressure drop across mill, in. water	11.1	11.4	11.2
Primary Cyclone Data:			
Inlet pressure N, in. water	−10.5	−10.7	−10.8
Outlet pressure N, in. water	−11.5	−11.7	−11.9
Inlet pressure S, in. water	−10.5	−10.7	−10.8
Outlet pressure S, in. water	−12.0	−12.1	−12.3
Pressure drop N, in. water	1.0	1.0	1.1
Pressure drop S, in. water	1.5	1.4	1.5
Main Exhauster Fan Data:			
Disch. temp, dry bulb, F	130	131	129
Relative humidity, per cent	36	38	36
Inlet press. before control damper, in. water	−12.1	−12.1	−12.3
Inlet press. after control damper, in. water	−27.2	−27.2	−27.2
Discharge pressure,[a] in. water	2.7	2.6	2.7
Discharge pressure,[b] in. water	3.6	3.6	3.6
Fan speed, rpm	1180	1180	1185
Air discharged, lb per hr	70800	72300	70800
Air discharged, cfm	18600	19000	18600
Cloth-Filter Data:			
Inlet temperature, F	130	131	129
Inlet-pipe velocity, fpm	3800	3880	3800
Inlet pressure,[a] in. water	2.3	2.3	2.3
Inlet pressure,[b] in. water	3.1	3.0	3.1
Coal collected, weighed, lb per hr	440	458	447
Outlet temperature, F	124	123	123
Dust concentration, gr per cu ft air	2.8	2.8	2.8
Collected coal, per cent of milled coal	0.5	0.6	0.6
Coal-separation eff of primary cyclones, per cent	99.5	99.4	99.4

[a] Neutral period (all filter compartments filtering).
[b] Cleaning period (one filter compartment closed for shaking).

shown is that it permits continuous movement of the coal dust in the pipe line, because it incorporates a storage hopper in addition to the weighing hopper and also that it is dust-tight which, in this particular application, is rather important. It is, of course,

Mill Heat Balance:			
Input			
With hot air, M Btu per hr	2590	2440	2390
Mill motor power, M Btu per hr	1120	936	961
Total, M Btu per hr	3710	3376	3351
Output			
Sensible heat in coal, M Btu per hr	656	648	606
Heat-up and evap moisture, M Btu per hr	1861	2013	2007
Heat-up inleakage air, M Btu per hr	13	20	27
Heat in air leaving mill, M Btu per hr	596	457	333
Total accounted for, M Btu per hr	3126	3138	2973
Rad. and unaccounted for, M Btu per hr	584	238	378
Expressed in per cent:			
Sensible heat in coal	17.7	19.2	18.1
Heat-up and evap moisture	50.2	59.6	59.9
Heat-up inleakage air	0.4	0.6	0.8
Heat in air leaving mill	16.0	13.5	9.9
Total accounted for	84.3	92.9	88.7
Rad. and unaccounted for	15.7	7.1	11.3
Total	100.0	100.0	100.0
Moisture Removal:			
Calculated			
Total in air from mill, lb per hr	2480	2620	2570
Less moisture in inlet air, lb per hr	779	846	870
Moisture removed, lb per hr	1701	1774	1700
Total moisture removed by analysis, lb per hr	1610	1340	1260
Moisture removed, calculated, per cent	2.0	2.3	2.3
Moisture removed, by analysis, per cent	1.9	1.8	1.7

Raw- and Pulverized-Coal Data

Screenings:			
Raw coal (mill inlet)			
Over 5/16-in. opening, per cent	13.1	34.6	32.8
Over 3/16-in. opening, per cent	26.5	23.5	25.0
Over 3/32-in. opening, per cent	22.8	16.5	16.9
Over 1/16-in. opening, per cent	10.0	6.2	6.5
Over 3/64-in. opening, per cent	5.5	3.7	3.8
Dust through 3/64-in. opening, per cent	22.1	15.5	15.0
Over 40, per cent	10.6	7.9	7.0
Over 60, per cent	3.9	2.8	2.7
Over 80, per cent	1.6	1.0	1.0
Over 100, per cent	1.0	0.7	0.7
Over 200, per cent	1.9	1.3	1.5
Through 200, per cent	3.1	1.8	2.1
Powdered Coal (milled):			
Over 40-mesh Tyler screen, per cent	1.9	1.7	1.8
Over 60-mesh Tyler screen, per cent	4.9	5.1	5.0
Over 80-mesh Tyler screen, per cent	5.0	5.1	4.7
Over 100-mesh Tyler screen, per cent	5.8	6.1	5.9
Over 200-mesh Tyler screen, per cent	20.5	20.6	20.3
Through 200-mesh Tyler screen, per cent	61.9	61.4	62.3
Reclaimed Coal (collected by filter):			
Over 200-mesh Tyler screen, per cent	0.0	0.0	0.0
Over 325-mesh Tyler screen, per cent	0.1	0.0	0.0
Through 325-mesh Tyler screen, per cent	99.9	100.0	100.0
Moisture Analysis:			
Raw coal, per cent	3.0	3.0	2.9
Powdered coal, per cent	1.1	1.2	1.2
Removed by analysis, per cent	1.9	1.8	1.7
Coal returned by cloth filters, per cent	0.8	1.3	1.3

Raw Coal—Proximate Analysis:	Wet basis	Dry	Wet basis	Dry	Wet basis	Dry
Moisture, per cent	3.0		3.0		2.9	
Volatile, per cent	29.5	30.4	33.1	34.1	31.5	32.4
Ash, per cent	6.5	6.7	6.7	6.9	6.6	6.8
Fixed carbon, per cent	61.0	62.9	57.2	59.0	59.0	60.8
Heat content, Btu per lb	14060	14490	13735	14160	13885	14330

Powdered-Coal Densities, A.S.T.M. Method:			
Milled coal, lb per cu ft	33.6		
Coal returned by cloth filters, lb per cu ft	17.8		

important if the dust-conveying pipe is under pressure other than atmospheric that the weighing hopper be vented during the weighing period. For this purpose, a vent valve is provided in the weighing-hopper cover plate. This particular type of installation has also been used successfully in weighing flue dust from pulverized-fuel installations. The hoppers are fabricated from 3/32-in. plate and all seams are welded.

Discussion

J. E. FULWEILER.[7] It would be interesting to know if any serious objections to the installation discussed in this paper were offered by the engineers of the insurance company covering the risk; also whether the demonstrated safe operation of this completely dry equipment was satisfactory proof that sprays are not necessary when the other precautions described are taken.

In view of the tendency to insist upon sprays in connection with explosive industrial dusts, this example may prove of considerable value in the art, because some of the dusts involved are of value if kept dry and utterly ruined if wet.

The wetting of the dusts also seems to increase rather than minimize the disposal problem as well as to entail considerable trouble, due to the freezing of the water in cold weather. An installation of this size and importance, operating dry, is therefore of considerable interest to the writer.

RALPH A. SHERMAN.[8] The writer would like to suggest that supplementary information covering the proximate analysis and calorific value of the dust, which is collected in the bag filters, would aid in a comparison of the remainder of the coal collected in the cyclones.

R. F. THRONE.[9] There are upward of 100 installations of cloth filters in service, handling vents from pulverized-coal equipment. The general experience has been quite satisfactory. Some provisions have been necessary to drain off the static electrical charge, and some installations have required special attention to the slope of the inclined surfaces. Only a few of the installations use flue gas as an inert gas, the circulating gas in most cases being air.

Cloth-type filters have the highest reclaim capacity, especially in their ability to remove the microscopic sizes under 10μ. These sizes represent as high as 80 per cent of the total dust and constitute a major visual nuisance.

Precaution must be exercised to maintain the gas humidity below 100 per cent, as the resistance through the cloth greatly increases as it absorbs free moisture, as well as permits the coal dust to cake onto the filter surface, thus increasing the difficulty of restoring the filtering medium. The tubular free-hanging type of cloth filter, can be adequately cleaned in service merely by closing the outlet-gas damper without in turn closing the inlet-gas damper. The agitating mechanism is sufficient to shake the coal dust from the filter cloth without allowing a time period for equalization of gas pressure on either side. The most effective procedure is to have a reverse flow of the gas through the filter cloth during the shaking operation, but such reversal, either with the cleaned gas from other sections or with room air, cools the gas in the compartment being cleaned, to and below its dew point, causing condensation. Cloth filters have proved equally successful under negative pressures as high as 26 in., as well as under positive pressure. The major disadvantage of negative-pressure operation is the detection of air-inleakage areas.

Fire hazard is increased in the employment of cloth filters. This becomes an important factor, especially with the "younger" coals, for which the affinity for oxygen is high. The combustion of filtering cloth with the resulting flame adds a menacing factor to what would otherwise be an easily controllable condition. Flue gas as the circulating medium has not proved satisfactory

in avoiding ignition of coal dust and the resulting loss of filter bags on the "western" coals. Even though a noninflammable filtering cloth material should be developed, considerable question remains whether with these "younger" coals, the large area available is not a definite hazard.

J. C. WITT.[10] The writer would like to know whether any installations of the type described have been involved in fires? He has had some experience with one all-steel dust-collector installation in which a fire occurred. Much of the equipment was rendered worthless by the heat, although nothing was burned. The question therefore arises: What benefit would come from fire-proofing the cloth, in this instance, if a fire should start?

AUTHORS' CLOSURE

The authors believe that the successful operation of the cloth-filter installation described is in a large measure made possible by the safety features incorporated in the design and also because of the systematic inspection program carried out by the operating personnel. The insurance carrier, in the case of this installation, interposed no requirements with those on pulverized-coal installations in general.

The powdered-coal proximate analyses omitted in the performance tests and requested by Mr. Sherman have been obtained. Average laboratory results of the powdered-coal analyses are tabulated as follows:

POWDERED-COAL PROXIMATE ANALYSES

Milled coal		Reclaimed coal (collected by cloth filter)
Moisture, per cent	1.4	0.8
Volatile matter, dry, per cent	34.4	33.8
Ash, dry, per cent	7.7	7.4
Fixed carbon, dry, per cent	57.9	58.8
Sulphur, dry, per cent	1.45	1.10
Calorific value, dry basis, Btu per lb	14110	14150

No appreciable difference in laboratory analyses of the two powdered-coal samples is evident. Although the primary cyclone efficiency is better than 99 per cent, it is not sufficient to detect differences in proximate analysis between milled coal and the vented coal, if such a difference exists. For this installation at least, the conclusion must be that the proximate analyses of the powdered coals are the same.

Mr. Throne points out that it is necessary to insure against the hazards of static electric discharges. This insurance is largely inherent in the collector described. The importance of maintaining the temperature of the coal dust conveying air in the collector above the dew point cannot be overemphasized. Careful selection of heat-insulating materials has eliminated difficulties of this kind.

Pressure balance of the cloth may be secured by closing either the inlet- or outlet-gas damper. It is not necessary to close both. In this installation, the inlet dampers were used because it simplified construction and outlet dampers were not provided because they were unnecessary. It has been found by experience that shaking in a screen-type collector is ineffective if the cloth pressure is unbalanced as little as 0.02 in. of water. Reverse gas flow is unnecessary and is of doubtful value.

Mr. Throne states that cloth pressure balance is unnecessary with the tubular free-hanging-type collector. It is quite unnecessary with all types if draft loss is no object. It would be of interest to know the magnitude of the pressure unbalance for the installation mentioned. It is certain that, if a pressure unbalance exists at the time of shaking, it is also true that more effective shaking is possible if the pressure unbalance is eliminated. Cloth maintenance will be reduced as a result. Needlessly high

[7] Power and Industrial Engineer, Schmid & Fulweiler, Philadelphia, Pa. Mem. A.S.M.E.

[8] Supervisor, Fuels Division, Battelle Memorial Institute, Columbus, Ohio. Mem. A.S.M.E.

[9] Secretary-Treasurer, Public Service Company of Colorado, Denver, Colo. Mem. A.S.M.E.

[10] Technical Service Manager, Marquette Cement Manufacturing Company, Chicago, Ill. Mem. A.S.M.E.

pressure drops should be avoided on large installations because fan-power requirements will be high and cloth maintenance will be a very costly item. Although the collector described operated at a pressure above atmospheric, the system will operate equally satisfactorily at a pressure below atmospheric. The pressure-release system in the latter case will be rather simple.

The authors agree with Mr. Witt that fireproofing the cloth does not protect the dust-filter equipment after a fire has started in the collector. As pointed out in the paper, the primary purpose of the cloth flameproofing is that of localizing a fire after it has started. At this stage, the chief concern of the operator would be to prevent spreading to other parts of the mill system and to adjacent equipment. From this point of view as well as the moral value, the slight additional cost of the flameproofing is justified. The authors do not know of any coal-dust cloth-filter installations of the type of design described in which fires have occurred. There have been fires in the type of collector described with bakelite-dust installations and other similar dusts. None of the fires were directly attributed to the collector but they were carried into it from other parts of the system.

Mean Temperature Difference in Design

By R. A. BOWMAN,[1] A. C. MUELLER,[2] AND W. M. NAGLE[3]

In heat-transfer apparatus the rate of heat flow from the hot to the cold fluid is proportional to the temperature difference between the two. For design purposes, it is necessary to be able to determine the mean difference in temperature from the inlet and exit temperatures. Numerous investigators have contributed analyses of the temperature difference for exchangers with neither counternor cocurrent flow. This paper coordinates the results of previous studies on the same basis to give as complete a picture as possible of all the various arrangements of surface and flow. Shell-and-tube exchangers with any number of passes on shell side and tube side are covered as are the crossflow exchangers with different pass arrangements and with mixed and unmixed flow. The special cases of trombone coolers, pot coolers, and batch processes, not previously published, are also treated in detail.

NOMENCLATURE

THE following nomenclature is used in the paper:

A = area of heating surface
C = specific heat of shell-side (or hot) fluid
c = specific heat of tube-side (or cold) fluid
e = base of natural logarithms
F = correction factor, dimensionless
M = weight of fluid batch
N = number of shell-side passes
P = $(t_2 - t_1)/(T_1 - t_1)$, dimensionless
p = $(T_1 - T_2)/(T_1 - t_1)$, dimensionless
Q = quantity of heat transferred
q = $(t_2 - t_1)/(T_1 - t_1)$, dimensionless
R = wc/WC, or $(T_1 - T_2)/(t_2 - t_1)$, dimensionless
r = $\Delta t_m/(T_1 - t_1)$, dimensionless
r_0 = value of r for countercurrent flow
T_1 = inlet temperature shell-side (or hot) fluid
T_2 = outlet temperature shell-side (or hot) fluid
t_1 = inlet temperature tube-side (or cold) fluid
t_2 = outlet temperature tube-side (or cold) fluid
Δt_m = mean temperature difference
U = over-all heat-transfer coefficient
W = weight rate of flow of shell-side (or hot) fluid
w = weight rate of flow of tube-side (or cold) fluid
Θ = time

INTRODUCTION

Heat-exchanger design is based primarily on the equation

$$Q/A\Theta = U\Delta t_m \dots \dots \dots \dots \dots [1]$$

[1] Section Engineer, Heat Exchanger and Condenser Section, Westinghouse Electric & Manufacturing Company, Philadelphia, Pa. Mem. A.S.M.E.
[2] Chemical Engineer, Technical Division, Engineering Department, E. I. du Pont de Nemours & Company, Inc., Wilmington, Del.
[3] Chemical Engineer, Eastern Laboratory, Explosives Department, E. I. du Pont de Nemours & Company, Inc., Gibbstown, N. J.

In this basic equation, the amount of heat transferred per unit of time and surface is shown to equal the product of the over-all heat-transfer coefficient U, and the mean temperature difference, Δt_m. The value of U may be estimated from correlated over-all coefficients determined in similar exchangers under similar conditions, or it may be obtained by combination of individual coefficients. The mean temperature difference Δt_m is calculated from terminal temperatures, the method of calculation varying with the type of exchanger and method of operation.

For simple heat exchangers in which there is steady-state operation and countercurrent flow of the hot and cold fluids, integration of the differential equation, relating the temperatures of the two fluids, leads to the well-known log-mean temperature difference

$$\Delta t_m \text{ (countercurrent flow)} = \frac{(T_1 - t_2) - (T_2 - t_1)}{2.3 \log_{10} \dfrac{T_1 - t_2}{T_2 - t_1}} \dots [2]$$

The integration involves the following assumptions:

1 The over-all heat-transfer coefficient U is constant throughout the heat exchanger.
2 The rate of flow of each fluid is constant.
3 The specific heat of each fluid is constant.
4 There is no condensation of vapor or boiling of liquid in part of the exchanger. Condensation of pure saturated vapor or boiling of pure saturated liquid throughout the entire length of the heat exchanger, resulting in $T_1 = T_2$ or $t_1 = t_2$, does not affect the integration if the first assumption remains true.
5 Heat losses are negligible.

For heat exchangers in which there is steady-state operation and cocurrent flow, the corresponding integrated average temperature difference is

$$\Delta t_m \text{ (cocurrent flow)} = \frac{(T_1 - t_1) - (T_2 - t_2)}{2.3 \log_{10} \dfrac{T_1 - t_1}{T_2 - t_2}} \dots \dots [3]$$

For any set of terminal temperatures, the average temperature difference for cocurrent flow is always less than that for countercurrent flow, unless the temperature of one fluid stream is constant throughout the exchanger.

In the majority of industrial installations true countercurrent heat exchangers are not as economical as multipass and crossflow units. In multipass exchangers the flow is cocurrent in part and countercurrent in part and, as a result, the mean temperature difference lies somewhere between Δt_m for cocurrent flow and Δt_m for countercurrent flow. The mean temperature difference for the various types of crossflow exchangers is also less than that for countercurrent flow and greater than that for cocurrent flow. The relationships between terminal temperatures and mean temperature difference limit the performance of each type of exchanger. For a given set of temperature and rate conditions, the mean temperature difference for some types of multipass or crossflow exchangers may be zero, and the exchanger therefore inoperable, while a countercurrent or another type of multipass exchanger may operate satisfactorily. Formulas or curves are available for calculating the mean temperature difference from terminal temperatures in various types of heat exchangers. These are assembled in this

paper and, where possible, are expressed in the form of a correction factor, F, by which the log mean Δt for countercurrent flow is multiplied to give the true mean temperature difference.

$$\Delta t_m = (F)\ (\Delta t_m \text{ for countercurrent flow})$$

$$= (F)\ \frac{(T_1 - t_2) - (T_2 - t_1)}{2.3\ \log_{10} \dfrac{T_1 - t_2}{T_2 - t_1}} \quad \cdots\cdots\cdots\cdots [4]$$

This method of presentation is believed to show most clearly the degree to which the mean temperature difference of any exchanger is inferior to the log-mean Δt for countercurrent flow.

MULTIPASS HEAT EXCHANGERS

The widely used shell-and-tube heat exchangers are often provided with baffled heads to route the fluid inside the tubes back and forth from one end of the exchanger to the other. In some cases the shell-side fluid is also caused to travel the length of the heat exchanger more than once by means of longitudinal baffles. Differential equations for a number of such arrangements have been derived and integrated. In the integration the following assumptions are made:

1　The over-all heat-transfer coefficient U is constant throughout the heat exchanger.
2　The rate of flow of each fluid is constant.
3　The specific heat of each fluid is constant.
4　There is no condensation of vapor or boiling of liquid in part of the exchanger.
5　Heat losses are negligible.
6　There is equal heat-transfer surface in each pass.
7　The temperature of the shell-side fluid in any shell-side pass is uniform over any cross section.

The first five assumptions are those employed in the derivation of the ordinary log-mean temperature-difference formula. The sixth is in accord with usual heat-exchanger design practice and the seventh is essentially true where many transverse baffles are incorporated in the exchanger.

Multipass exchangers with an even number of tube-side passes per shell-side pass were studied by Nagle (1),[4] Underwood (2), Bowman (3), and Yendall(15). Their results are summarized as follows:

One Pass Shell Side; Two Passes Tube Side. The correction factor F for multipass heat exchangers, having one shell-side and two tube-side passes, is plotted in Fig. 1 against P and R. The latter are dimensionless ratios, defined as follows:

---

[4] Numbers in parentheses refer to the Bibliography at the end of the paper.

$$P = (t_2 - t_1)/(T_1 - t_1)$$

$$R = (T_1 - T_2)/(t_2 - t_1)$$

The curves of Fig. 1 are based on the integrated equations obtained by Underwood (2) and by Yendall, rather than on the graphical integrations of Nagle (1), from which they differ by as much as 2 per cent. The integrated equation for Δt_m is

Δt_m (one-two exchanger) =

$$\frac{\sqrt{(T_1 - T_2)^2 + (t_2 - t_1)^2}}{\log_e \left\{ \dfrac{T_1 + T_2 - t_1 - t_2 + \sqrt{(T_1 - T_2)^2 + (t_2 - t_1)^2}}{T_1 + T_2 - t_1 - t_2 - \sqrt{(T_1 - T_2)^2 + (t_2 - t_1)^2}} \right\}} \quad \cdots [5]$$

This equation may be transposed into the following form, in which the F factor for one-two exchangers is expressed in terms of P and R.

$$F_{1,2} =$$

$$\frac{\sqrt{R^2 + 1}}{R - 1} \log_{10} \frac{1 - P}{1 - PR} \bigg/ \log_{10} \frac{(2/P) - 1 - R + \sqrt{R^2 + 1}}{(2/P) - 1 - R - \sqrt{R^2 + 1}}$$

$$\cdots [6]$$

In Equation [6], and in subsequent Equations [7] and [8], the expression

$$\frac{1}{R - 1} \log_{10} \frac{1 - P}{1 - PR}$$

becomes indeterminate when $R = 1$, but the usual treatment for such indeterminates reduces this expression to

$$\frac{P}{2.3\ (1 - P)}$$

The correction factor is exactly the same whether the shell-side fluid enters at the fixed or the floating head.

One Pass Shell Side; Four Passes Tube Side. The correction factor for one-four heat exchangers is slightly less than that for one-two exchangers, but the difference is so small that separate curves are unnecessary. Underwood's integration is expressed in hyperbolic functions, and Yendall's in logarithms; but the numerical results appear to be the same. Based on Yendall's work, the correction factor may be expressed as

$$F_{1,4} =$$

$$\frac{\sqrt{4R^2 + 1}}{2(R - 1)} \log \frac{1 - P}{1 - PR} \bigg/ \log \frac{1 + V\left(\sqrt{4R^2 + 1} - 2R\right)}{1 - V\left(\sqrt{4R^2 + 1} + 2R\right)} \cdots [7]$$

FIG. 1　CORRECTION-FACTOR PLOT FOR EXCHANGER WITH ONE SHELL PASS AND TWO, FOUR, OR ANY MULTIPLE OF TUBE PASSES

FIG. 2 CORRECTION-FACTOR PLOT FOR EXCHANGER WITH TWO SHELL PASSES, AND FOUR, EIGHT, OR ANY MULTIPLE OF TUBE PASSES

FIG. 3 CORRECTION-FACTOR PLOT FOR EXCHANGER WITH THREE SHELL PASSES, AND SIX, TWELVE, OR ANY MULTIPLE OF TUBE PASSES

FIG. 4 CORRECTION-FACTOR PLOT FOR EXCHANGER WITH FOUR SHELL PASSES, AND EIGHT, SIXTEEN, OR ANY MULTIPLE OF TUBE PASSES
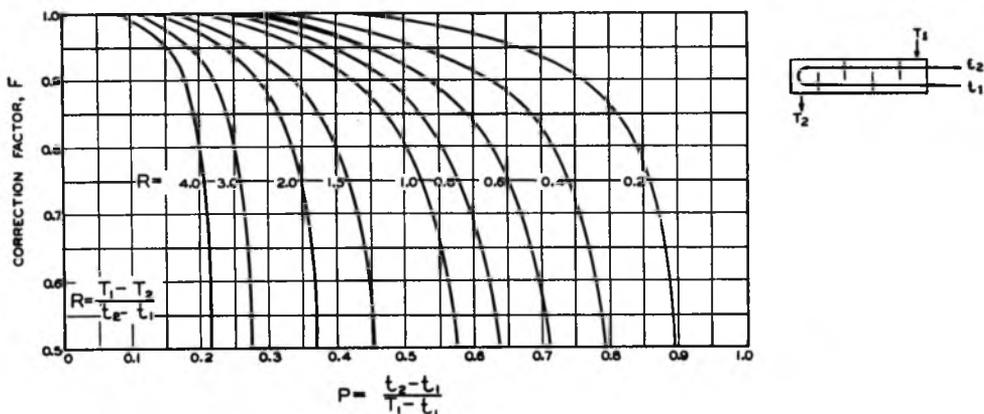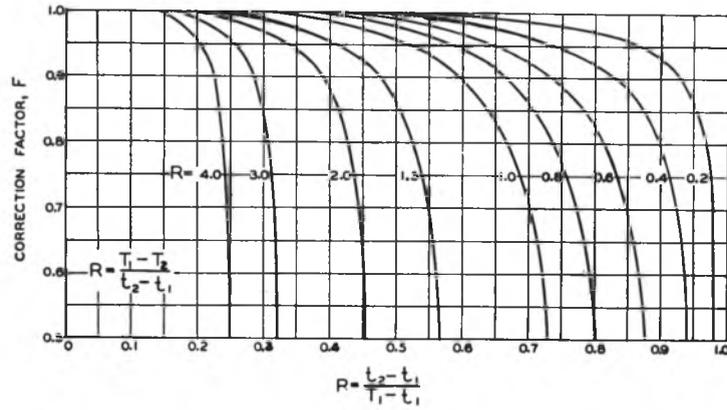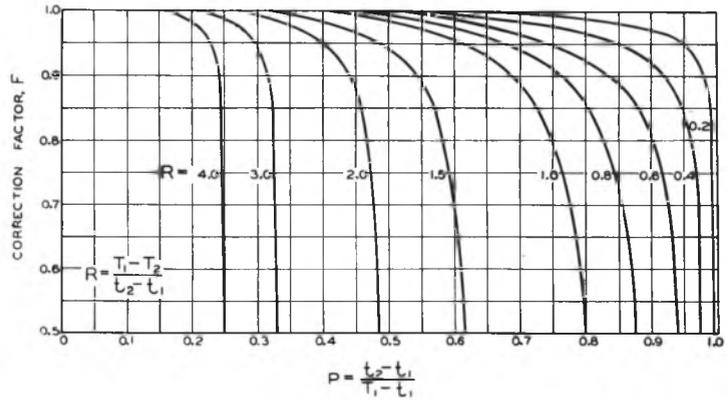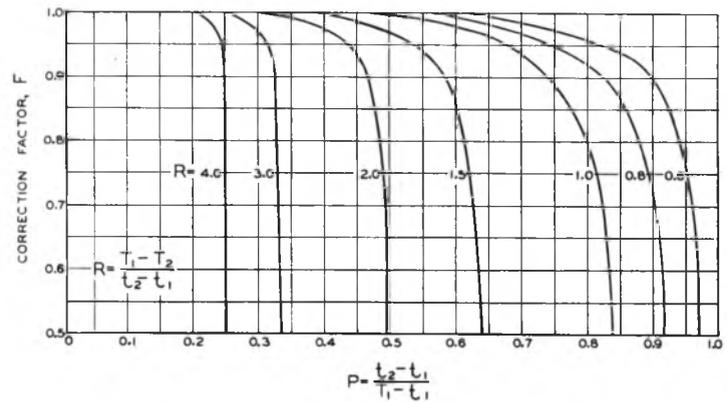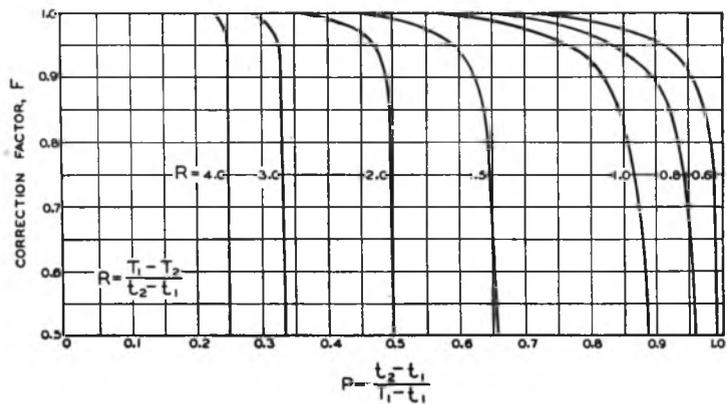

FIG. 5 CORRECTION-FACTOR PLOT FOR EXCHANGER WITH SIX SHELL PASSES, AND TWELVE, TWENTY-FOUR, OR ANY MULTIPLE OF TUBE PASSES
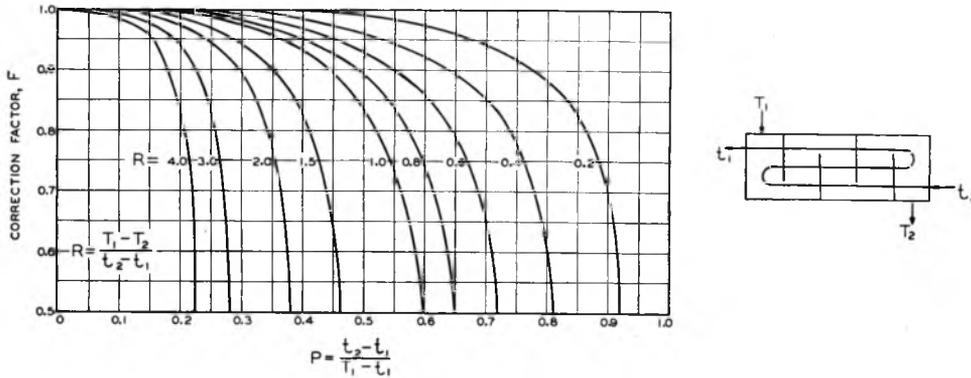
FIG. 6   CORRECTION-FACTOR PLOT FOR EXCHANGER WITH ONE SHELL PASS, AND THREE, SIX, OR ANY MULTIPLE OF TUBE PASSES;
MAJORITY OF TUBE PASSES IN COUNTERCURRENT FLOW

where

$$V = \frac{t_2 - t_1}{4T_1 - (t_1 + 2t_i + t_2)}$$

and t_i is the intermediate temperature of the tube-side fluid where it leaves the second and enters the third tube-side pass. The value of t_i may be obtained by trial-and-error solution of

$$\left(\frac{t_i - t_1}{t_2 - t_i}\right)^{\sqrt{4R^2+1}} = \frac{1 + V\left(\sqrt{4R^2 + 1} - 2R\right)}{1 - V\left(\sqrt{4R^2 + 1} + 2R\right)}$$

One Pass Shell Side; Six Passes Tube Side. Nagle (1) found that the correction factors for one-six exchangers agreed so closely with those of one-two exchangers operating with the same terminal temperatures that the curves of Fig. 1 may be used for this case.

One Pass Shell Side; Infinite Passes Tube Side. It has been pointed out by Bowman (3) that the correction factor F for a multipass heat exchanger, having one shell-side pass and a very large number of tube-side passes, approaches as a limit the case of crossflow with both fluids completely mixed. Even in the limit, however, the value of F is generally only 1 to 2 per cent less than that of the one-two exchangers.

Two Passes Shell Side; Four Passes Tube Side. The foregoing discussion indicates that the correction factors for heat exchangers with one shell-side pass and two, four, six, and infinite tube-side passes are essentially the same. In many jobs that reach the designer, however, these types of exchangers are inefficient or entirely inoperative, and an arrangement giving higher values of F is required. This may be obtained by the use of two exchangers in series or of a single exchanger having two shell-side passes and four tube-side passes. The correction factor F for two-four exchangers piped in the usual manner to approach countercurrent flow is given graphically in Fig. 2 and algebraically by the following rearrangement of Underwood's equation for two-four heat exchangers

$$F_{2,4} = \frac{\sqrt{R^2 + 1}}{2(R - 1)} \log_{10} \frac{1 - P}{1 - PR}$$

$$\Big/ \log_{10} \frac{(2/P) - 1 - R + \frac{2}{P}\sqrt{(1 - P)(1 - PR)} + \sqrt{R^2 + 1}}{(2/P) - 1 - R + \frac{2}{P}\sqrt{(1 - P)(1 - PR)} - \sqrt{R^2 + 1}}$$

$$\dots\dots[8]$$

The derivations leading to Fig. 2 and Equation [8] involve the additional assumptions that there is no leakage of fluid or heat across the transverse baffle separating the two shell-side passes.

Three and More Shell-Side Passes. Bowman (3) developed a general method for calculating the F factors of three-six, four-eight, six-twelve, etc., heat exchangers from the correction factors of one-two exchangers. At any given values of F and R, the value of P for an exchanger having N shell-side and $2N$ tube-side passes is related to P for a one-two exchanger by the equation

$$P_{N,2N} = \frac{1 - \left(\dfrac{1 - P_{1,2}R}{1 - P_{1,2}}\right)^N}{R - \left(\dfrac{1 - P_{1,2}R}{1 - P_{1,2}}\right)^N} \dots\dots\dots\dots[9]$$

For the special case of $R = 1$, Equation [9] becomes an indeterminate which reduces to

$$P_{N,2N} = \frac{P_{1,2}N}{P_{1,2}N - P_{1,2} + 1}$$

The results are given in Figs. 3, 4, and 5, which may also be used with little error for exchangers that are multiples of the one-four or the one-six, rather than the one-two exchanger.

From an inspection of Figs. 1 through 5, it may be seen that, for any value of P and R, the correction factor approaches unity as the number of shell-side passes is increased. This is to be expected, since a multipass exchanger with several shell-side passes approaches the ideal countercurrent heat exchanger more closely than does one with one shell-side and two or more tube-side passes.

In all the cases mentioned, the ratio of tube-side to shell-side passes is an even number, such as 2, 4, or 6. However, Fischer (4) has pointed out that there is some improvement in the mean temperature difference if the ratio of tube-side to shell-side passes is an odd number and the exchanger is so connected that the tube-side fluid is flowing counter to the shell-side fluid in over half the passes.

One Pass Shell Side; Three Passes Tube Side. The correction factors F, for one-three heat exchangers and for two-six, three-nine, and four-twelve exchangers have been derived by Fischer (4) and presented in the form of tables and charts. In each case, the value of F is greater than is found in exchangers having the same number of shell-side passes but an even numbered ratio of tube to shell passes. The improvement in mean temperature difference resulting from the use of three tube-side passes per shell-side pass instead of two or four is, however, by no means as great as that resulting from an increase of one in the number of shell-side passes. Correction factors for the one-three exchanger are given in Fig. 6.

Effect of Variation in Heat-Transfer Coefficient. The preced-

ing derivations are based on the assumption that the over-all heat-transfer coefficient U is constant throughout the heat exchanger. For the case of countercurrent flow in which U varies linearly with temperature, Colburn (13) has derived the general heat-transfer equation

$$\frac{Q}{A\theta} = \frac{U_2 (T_1 - t_2) - U_1 (T_2 - t_1)}{2.3 \log_{10} \dfrac{U_2 (T_1 - t_2)}{U_1 (T_2 - t_1)}}$$

where U_1 is the value of U at the T_1 end of the exchanger and U_2 at the T_2 end. This equation has been combined by Sieder and Tate (*Industrial and Engineering Chemistry*, vol. 28, 1936, p. 1434) with the correction factor F for use with multipass heat exchangers

$$\frac{Q}{A\theta} = (F) \frac{U_2 (T_1 - t_2) - U_1 (T_2 - t_1)}{2.3 \log_{10} \dfrac{U_2 (T_1 - t_2)}{U_1 (T_2 - t_1)}}$$

Until this equation has been tested more extensively its use should be limited to multipass exchangers having only one shell-side pass.

EXAMPLES APPLYING TO MULTIPASS EXCHANGERS

The examples illustrating the variation of the mean temperature difference with type of heat exchanger are shown in Table 1.

TABLE 1 EXAMPLES ILLUSTRATING VARIATION OF MEAN TEMPERATURE DIFFERENCE WITH TYPE OF HEAT EXCHANGER

Example No. 1:
Where $T_1 = 300$ F, $T_2 = 200$ F, $t_1 = 100$ F, $t_2 = 200$ F
$(R = 1, P = 0.5)$

Exchanger	F	Δt_m, deg F
Countercurrent flow	1.000	100.0
Six-twelve multipass	1.000	100.0
Two-pass countercurrent crossflow, one fluid mixed	0.970	97.0
Horizontal film-type, helical connection	0.960	96.0
Two-four multipass	0.958	95.8
Horizontal film-type, return-bend connection	0.915	91.5
Single-pass crossflow, both fluids unmixed	0.910	91.0
One-three multipass	0.840	84.0
Single-pass crossflow, one fluid mixed	0.835	83.5
One-two multipass	0.804	80.4
One-four multipass	0.798	79.8
Single-pass crossflow, both fluids mixed	0.790	79.0
Two-pass cocurrent crossflow, one fluid mixed	0.405	40.5
Cocurrent flow	0	0

Example No. 2:
Where $T_1 = 300$ F, $T_2 = 200$ F, $t_1 = 160$ F, $t_2 = 260$ F
$(R = 1, P = 0.714)$

Exchanger	F	Δt_m, deg F
Countercurrent flow	1.000	40.0
Six-twelve multipass	0.970	38.8
Horizontal film-type, helical connection	0.795	31.8
Two-pass countercurrent crossflow, one fluid mixed	0.745	29.8
Single-pass crossflow, both fluids unmixed	0.682	27.3
Two-four multipass	0.645	25.8
Horizontal film-type, return-bend connection	0.395	15.8
One-three multipass	Impossible	
Single-pass crossflow, one fluid mixed	Impossible	
One-two multipass	Impossible	
Single-pass crossflow, both fluids mixed	Impossible	
One-four multipass	Impossible	
Two-pass cocurrent crossflow, one fluid mixed	Impossible	
Cocurrent flow	Impossible	

In the first example, all arrangements except that of cocurrent flow can be used. In the second example only countercurrent-flow heat exchangers, or multipass exchangers, which have two or more shell-side passes, can possibly give the desired results.

In the design of multipass heat exchangers, the general rules of Fischer (4) applying to mean temperature difference should be kept in mind. These are quoted as follows:

1 When the number of tube passes per shell pass is even, the mean temperature difference is independent of the direction of flow of the shell fluid in each shell pass.

2 When the number of tube passes per shell pass is odd, the mean temperature difference depends on the direction of flow of the shell fluid in each shell pass and is greatest when the counter-flow tube passes exceed the parallel-flow tube passes in each shell pass.

3 In multi-shell-pass exchangers the over-all mean temperature difference depends on whether the flow of the fluids between the shell passes is counter or parallel flow.

4 The equations available for solving the true mean temperature difference applying to multipass heat exchangers can be made general by letting the T values represent the hot-fluid and t values represent the cold-fluid terminal temperature.

Use of Correction Factors in Design. From the standpoint of mean temperature difference alone, it is desirable to use the same number of shell passes as of tube passes since by so doing true countercurrent flow can be obtained. However, structural and servicing considerations limit the number of shell passes, two passes in general being the maximum number for one shell. On the other hand the minimum number of tube passes is set by the quantity of fluid, the allowable pressure drop, the available space, and the costs of construction. Any reduction in the number of tube passes below this figure will result in a lowered heat-transfer rate which may overbalance any gain in mean temperature difference.

The usual procedure in designing a heat exchanger is, then, to choose the number of tube passes on the basis of the foregoing considerations and then to select the number of shell passes which will give a satisfactory mean temperature difference. If the use of two passes gives too low a correction factor, the usual practice is to divide the required surface among two or more shells in series.

Probably the most noticeable feature of the correction curves is the fact that they all have a very steep gradient for the lower values of the correction factor. In other words, a relatively small change in temperature conditions will cause a large change of F and consequently of the mean temperature difference. It also follows that a small deviation from the assumed conditions on which the curves are based will cause an appreciable error in mean temperature differences derived from this part of the curves. For instance, an exchanger might be expected to work with a mean temperature difference of 50 per cent that of a countercurrent exchanger on the basis of the correction-factor curves, whereas, it may actually be an impossible condition, due to the fact that there is leakage of fluid or heat through the longitudinal shell baffle contrary to the assumption of zero leakage.

Consequently, the use of correction factors below 0.8 seems to be questionable from a design standpoint. In those cases where such low correction factors are indicated, one or more additional shell passes should be used. Also test data, obtained under conditions which give low correction factors, should not be accepted too readily for general application.

The use of exchangers with three tube passes per shell pass gives some improvement in mean temperature difference and there are applications where this is of value. Unfortunately, the greatest improvement comes at the lower values of correction factor where the use of the calculated data for any type of exchanger is none too safe. Furthermore, mechanical and thermal difficulties may prevent the use of this arrangement in the majority of cases.

When only the inlet temperatures, rates of flow, heat-transfer area, and coefficient are known, the method of Ten Broeck (5) is especially useful. The exit temperatures are easily obtained from his type of plot, eliminating the necessity of trial-and-error solution.
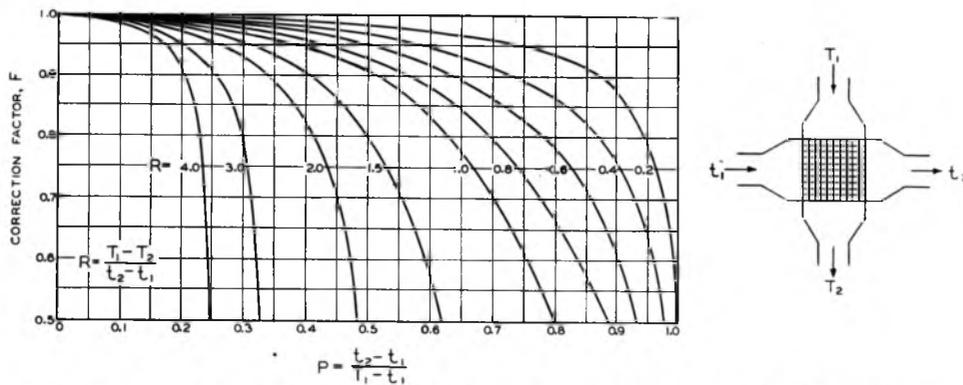
FIG. 7   CORRECTION-FACTOR PLOT FOR SINGLE-PASS CROSSFLOW EXCHANGER, BOTH FLUIDS UNMIXED

CROSSFLOW EXCHANGERS

Crossflow exchangers have the fluids flowing at right angles, therefore, neither the cocurrent nor countercurrent equations for mean temperature differences can be applied. The temperature of the fluid may vary in a direction normal to its flow as well as with the direction of flow. At the outlet, all infinitesimal sections of the stream may be at different temperatures as each was subjected to a different range of temperatures of the other fluid. The final outlet temperature is obtained by complete mixing of all infinitesimal sections. Mixing may take place within the exchanger and, if completely mixed, the fluid will have a temperature gradient only in the direction of flow. Therefore, the mean temperature difference is dependent upon whether either or both fluids are unmixed or mixed. In any practical crossflow exchanger, the case of unmixed fluids is more nearly approached than that of completely mixed fluids. Shell-and-tube exchangers with many transverse baffles are not crossflow exchangers because, even though the flow is normal to the tubes, the temperature change of the shell fluid is negligible for any baffle pass, the temperature gradient being parallel to the tubes. Conversely, as the number of shell-side passes in crossflow exchange is increased, the more closely will the case of cocurrent or countercurrent flow be approached.

Mathematical analysis of the mean temperature difference in a single-pass crossflow exchanger was initiated by Nusselt (6, 7). Smith (8) finished the analysis of the remaining cases of single-pass crossflow and extended it to two cases of two-pass crossflow exchangers. While others have presented equations or methods to determine the mean temperature difference, only Nusselt (7) and Smith (8) have carried out their calculations and presented tables or graphs from which the mean temperature difference may easily be computed from the known terminal temperatures.

There are several methods of presenting the results of the analyses of crossflow; each has certain advantages. Smith and Nusselt use a method in which the temperature-difference relationships depend upon three parameters which are defined as

$$p = \frac{T_1 - T_2}{T_1 - t_1}$$

$$q = \frac{t_2 - t_1}{T_1 - t_1}$$

$$r = \frac{\Delta t_m}{T_1 - t_1}$$

Substituting these parameters in the log-mean temperature-difference equation for countercurrent flow yields

$$r_0 \text{ (countercurrent flow)} = \frac{p - q}{\log_e \frac{1 - q}{1 - p}}$$

The method adopted for this paper is that of comparison to the log-mean temperature difference for countercurrent flow, which represents the maximum obtainable difference. This involves the use of F, P, and R which can be expressed in terms of the three parameters by the following relationships

$$F = \frac{r \text{ (crossflow)}}{r_0 \text{ (countercurrent flow)}}$$

$$R = \frac{p}{q} = \frac{T_1 - T_2}{t_2 - t_1}$$

$$P = q = \frac{t_2 - t_1}{T_1 - t_1}$$

By means of these relationships Smith's data (8) have been transposed into the curves Figs. 7 to 11. In the present form, the efficiency of the type of exchanger being considered is given by F. It is also easily apparent whether the exchanger will operate under the given conditions and what penalty is paid in the form of a poor correction factor F. For performance calculations on a known exchanger, the method of Ten Broeck (5) is especially useful.

The single-pass crossflow exchanger is the only case which has been completely investigated for all types of flow. Only several of the more important types of flow have been presented for the double-pass exchanger and the double-pass trombone exchanger. Mean temperature differences for more than two passes have not been determined except for the trombone exchangers. As the number of passes increases, the log-mean temperature difference for cocurrent or countercurrent flow will be approached, but no rule can now be given for the number of passes required before serious errors are introduced by use of these limits.

The integrations for mean temperature difference were made under the following assumptions:

1   The over-all heat-transfer coefficient U is constant throughout the heat exchanger.

2   The rate of flow of each fluid is constant.

3   The specific heat of each fluid is constant throughout the heat exchanger.

4   There is no condensation of vapor or boiling of liquid in part of the exchanger.

5   Heat losses are negligible.

6   There is equal heat-transfer surface in each pass.

7   The fluid is either unmixed or completely mixed normal to the flow, depending upon the type of flow assumed.

FIG. 8 CORRECTION-FACTOR PLOT FOR SINGLE-PASS CROSSFLOW EXCHANGER, ONE FLUID MIXED, OTHER UNMIXED

FIG. 9 CORRECTION-FACTOR PLOT FOR SINGLE-PASS CROSSFLOW EXCHANGER, BOTH FLUIDS MIXED

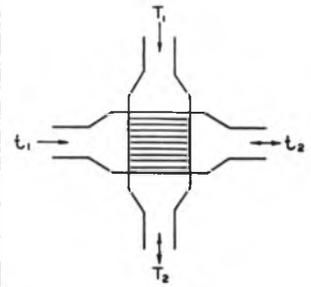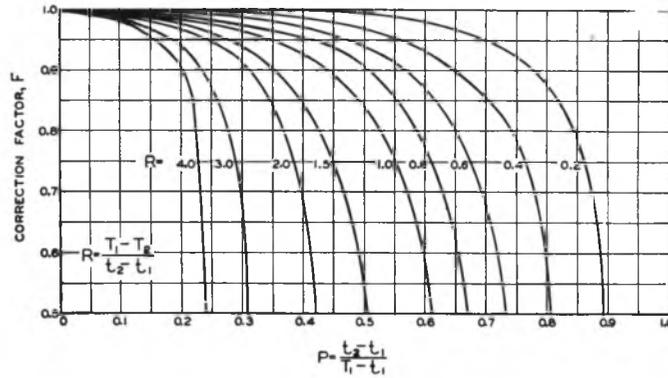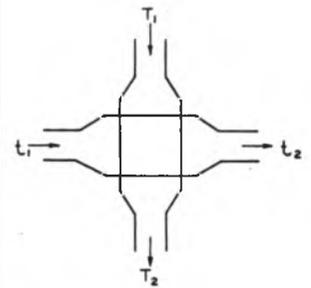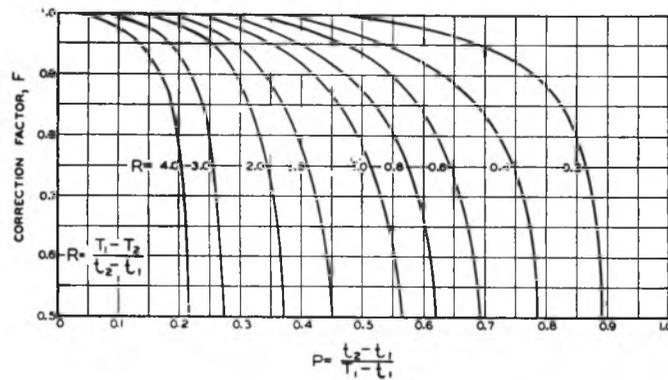
FIG. 10 CORRECTION-FACTOR PLOT FOR TWO-PASS CROSSFLOW EXCHANGER, SHELL FLUID MIXED, TUBE FLUID UNMIXED
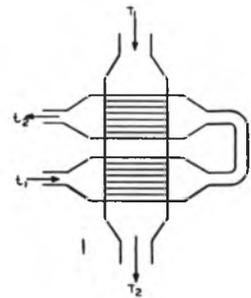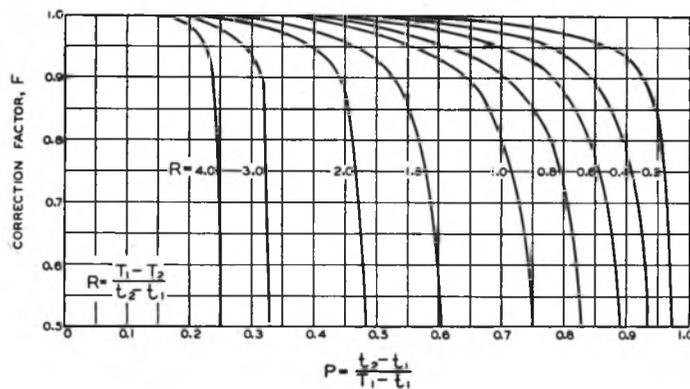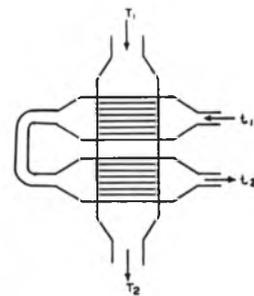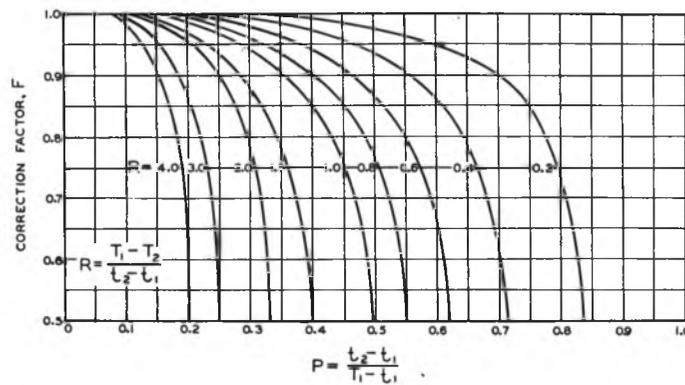(Shell fluid flowing across second and first passes in series.)

FIG. 11 CORRECTION-FACTOR PLOT FOR TWO-PASS CROSSFLOW EXCHANGER, SHELL FLUID MIXED, TUBE FLUID UNMIXED
(Shell fluid flowing across first and second pass in series.)

Single Pass; Both Fluids Unmixed. Each fluid crosses in a series of independent infinitesimal streams between which there is no mixing or heat transfer until the outlet is reached. Since each stream is subjected to a different range of temperatures of the other fluid, the streams will have different outlet temperatures, the final outlet temperature being the result of mixing of all streams. This assumption represents only the limit for any actual heat exchanger. However, it is nearly approached when one fluid flows through many tubes and the second fluid, crossing outside the tubes, is divided into many streams by baffles. In plate exchangers, mixing is prevented by baffles on each side.

This type of crossflow has been frequently investigated. Nusselt's first solution (6) contained slowly converging series and in a later paper (7) a more rapidly converging series was obtained. Roszak, Véron, and Tripier (9), Binnie and Poole (10), and Liebaut (11) have all presented solutions of varying complexity. Fig. 7 was obtained from values calculated by Nusselt and based on the equation

$$r = \sum_{u=0}^{u=\infty} \sum_{v=0}^{v=\infty} \left\{ (-1)^{u+v} \frac{(u+v)!}{u!(u+1)!v!(v+1)!} \left(\frac{p}{r}\right)^u \left(\frac{q}{r}\right)^v \right\} \quad \dots [10]$$

Single Pass; One Fluid Mixed, Other Unmixed. In this case, one fluid is so completely mixed normal to its flow that there is no temperature gradient. The second fluid is assumed to consist of independent streams between which there is no mixing. This is approached by the first fluid flowing across tubes with no baffles to prevent mixing and the second fluid flowing through tubes. This type of flow is obtained in a single-pass trombone exchanger. Here, as in subsequent cases where one fluid is mixed, p refers to the mixed fluid, q to the unmixed. Fig. 8 is based on the graphs of Smith (8) which were calculated from the equation

$$r = \frac{q}{\log_e \cfrac{1}{1 - \cfrac{q}{p} \log_e \cfrac{1}{1-p}}} \quad \dots \dots \dots [11]$$

Comparison of the graphs shows the correction factors F are lower than in the preceding case. If an exchanger is to be designed with predetermined terminal temperatures, but the choice of fluid which is to be mixed is open, a larger correction factor F is obtained if the fluid with the greater temperature change is chosen for the mixed fluid. However, the increase in most cases is small.

Single Pass; Both Fluids Mixed. The remaining limiting case for single-pass crossflow is where each fluid is completely mixed so there is no temperature gradient normal to its flow. In Fig. 9 the correction factor F, is lower than in either of the two other cases. Here it is possible to get into a region where reversed heat flow may take place, that is, near the outlet the original cold fluid may be hotter than the other fluid. Fig. 9 is based on the graphs of Smith (8) which were calculated from the equation

$$r \left\{ \frac{\dfrac{p}{r}}{1 - e^{-\frac{p}{r}}} + \frac{\dfrac{q}{r}}{1 - e^{-\frac{q}{r}}} - 1 \right\} = 1 \dots \dots [12]$$

Two-Pass Crossflow Exchangers. Two identical crossflow exchangers may be connected in series in at least ten different ways, as shown schematically in Table 2. Each of these arrangements can be operated either co- or countercurrent. The cases which have been partly or completely solved follow:

Ia Both Fluids Mixed. This case can be solved by trial

TABLE 2— SCHEMATIC ARRANGEMENTS OF VARIOUS CASES OF TWO-PASS CROSS-FLOW (CASES MARKED WITH TRIANGLE ARE EITHER PARTIALLY OR WHOLLY SOLVED)
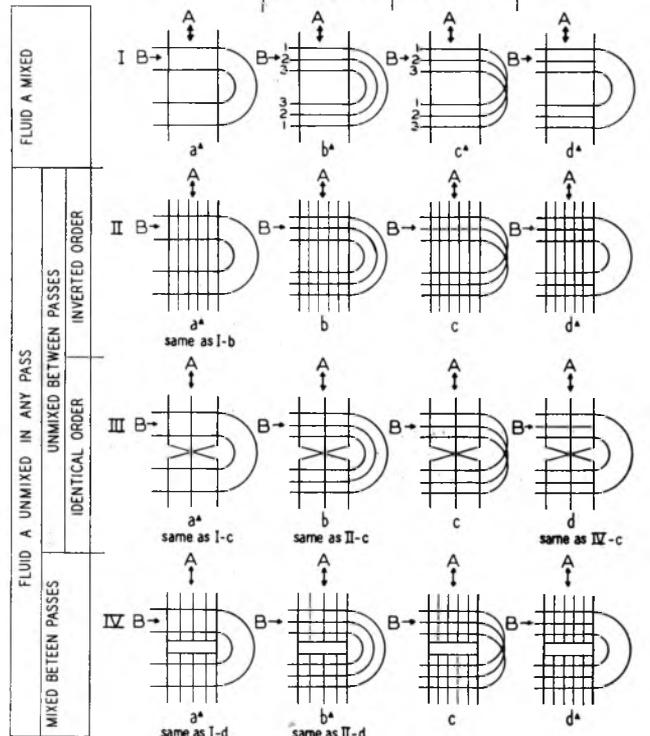
and error for the intermediate temperatures using the correction factors for the corresponding single-pass case. The correction factor for each pass is the same. Equation [9] is useful for countercurrent flow.

Ib Fluid A Mixed, Fluid B Unmixed, Passes Connected in Inverted Order. This is exemplified by the trombone cooler with return-bend connections.

Ic Fluid A Mixed, Fluid B Unmixed, Passes Connected in Identical Order. This is exemplified by the trombone cooler with helical connections.

Id Fluid A Mixed, Fluid B Unmixed, Except Between Passes.

Here it is assumed that the unmixed fluid makes two passes through the tubes but is mixed to a uniform temperature between the passes. The second fluid is completely mixed at all times and flows across the tubes of the second and first tube passes in series. Fig. 10 was obtained from the graphs by Smith (8) which were determined from the equation for the countercurrent case

$$r = \frac{q}{2 \log_e \cfrac{1}{1 - \cfrac{q}{p} \log_e \cfrac{\sqrt{\cfrac{1-q}{1-p}} - \cfrac{q}{p}}{1 - \cfrac{q}{p}}}} \quad \dots \dots [13]$$

When the present case is changed to cocurrent flow it is possible to obtain reversed heat flow due to the temperatures of the fluids crossing. The correction factors F in Fig. 11 apply to the log-mean temperature difference for countercurrent flow

as this is the basis for all other graphs. Fig. 11 is based on the graphs of Smith (8) which were calculated from the equation

$$r = \cfrac{q}{2 \log_e \cfrac{1}{1 - \cfrac{q}{p} \log_e \cfrac{p+q}{q+p} \sqrt{1-(p+q)}}} \quad \dots \dots [14]$$

IId Both Fluids Unmixed, One Mixed Between Passes. This case is treated by Schumann (14) for co- and countercurrent flow. The method of analysis developed depends upon calculating each pass separately. At the present time, no simple relationship has been developed to determine the mean temperature difference for both passes from the terminal temperatures. One pass is the same case developed by Nusselt, the other, however, is complicated by an initial temperature distribution entering the pass.

IVd Fluids A and B Unmixed But Mixed Between Passes. This is treated in the same manner as *Ia.*

HORIZONTAL FILM-TYPE EXCHANGERS

Horizontal film-type exchangers (also called trombone or trickle coolers) are simply banks of horizontal pipes, one above he other, over which a film of fluid is distributed. These exchangers are frequently used because of their simple and cheap construction and the ease of cleaning the outside of the pipe.

The equations for any number of passes of horizontal film-type exchangers have been derived. The cases solved are for countercurrent flow and for the return-bend and helical systems of connecting the passes. The usual assumptions given at the beginning of this section are made. In addition, it is assumed that the temperature of the fluid, inside the pipe, T, is uniform throughout its cross section at any point. It is also assumed that the liquor flowing over the banks of pipes is not mixed laterally.

The final temperature of the liquor, t_2, is the result of complete mixing of each independent stream, which was at a different temperature when leaving the bottom pipe.

Return Bends. Here the fluid in the tube in one pass flows in an opposite direction to the fluid in the pass immediately above or below. There is only one bank of pipes formed with this system of connections. Fig. 12 is based on curves obtained for two passes from the equation

$$\frac{1}{1-p} = e^{Kp/q} \left[\cosh \frac{Kp}{q} + (1-K) \sinh \frac{Kp}{q} \right] \dots [15]$$

where $K = 1 - e^{-q/2r}$.

Helical Connection. In this case the pipe may be considered as an elongated coil forming two banks of pipes. The direction of flow through all pipes in a given bank is the same. Fig. 13 is based on curves obtained for two passes from the equation

$$\frac{1}{1-p} = e^{Kp/q} \left(e^{Kp/q} - K^2 \frac{p}{q} \right) \dots \dots [16]$$

where $K = 1 - e^{-q/2r}$.

MEAN TEMPERATURE DIFFERENCE FOR TANKS IN SERIES, HEATED OR COOLED BY COILS

There are occasional processes where a fluid is heated or cooled in a series of tanks by coils. One fluid overflows from one tank to another, each tank being at a different temperature. Fluid in the coils may be in series countercurrent to the tank fluid or each coil in a tank may be fed separately. The problem in design or analysis of data arises when only the number of tanks and the terminal temperatures are known. A method has been derived for such cases. For this case, assumptions are made as follows:

FIG. 12 CORRECTION-FACTOR PLOT FOR TWO-PASS HORIZONTAL FILM-TYPE EXCHANGER, RETURN-BEND CONNECTIONS

$$R = \frac{T_1 - T_2}{t_2 - t_1}$$
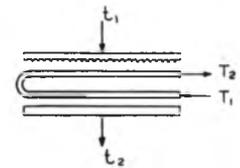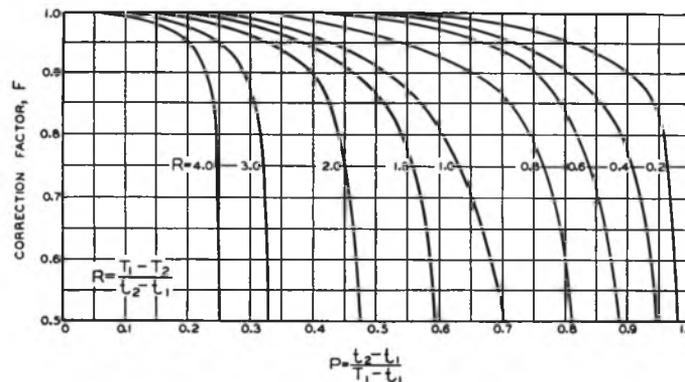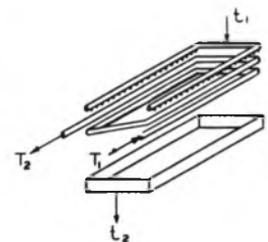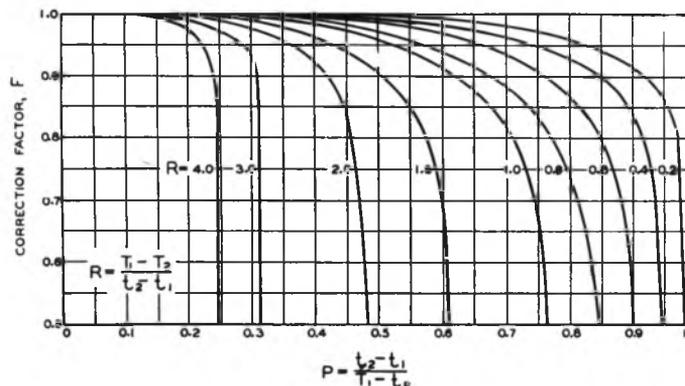
$$P = \frac{t_2 - t_1}{T_1 - t_1}$$

FIG. 13 CORRECTION-FACTOR PLOT FOR TWO-PASS HORIZONTAL FILM-TYPE EXCHANGER, HELICAL CONNECTIONS

$$R = \frac{T_1 - T_2}{t_2 - t_1}$$

$$P = \frac{t_2 - t_1}{T_1 - t_1}$$

1 Heat-transfer area and coefficient in each tank are the same.
2 Each tank is at a constant and uniform temperature.
3 Heat transferred is sensible heat.
4 Specific heats are constant.
5 Temperature and flow of fluids are constant.

Tanks in Series; Coils in Series Counterflow

Let $T_1, T_2, \ldots, T_n$ = temperature of feed to first, second, $\ldots n$ tank
           = temperature of coil liquid leaving the first, second, $\ldots n$ tank
           = temperature of liquid entering n tank

By a heat balance

$$\frac{wc}{WC} = R = \frac{T_n - T_{n+1}}{t_n - t_{n+1}} = \frac{T_1 - T_{n+1}}{t_1 - t_{n+1}} \ldots \ldots [17]$$

for any tank, say tank No. 2

$$wc\,(t_2 - t_3) = UA\, \frac{(T_3 - t_3) - (T_3 - t_2)}{\log_e \frac{T_3 - t_3}{T_3 - t_2}} \ldots [18]$$

and with UA constant for all tanks

$$\frac{UA}{wc} = \log_e \frac{T_{n+1} - t_{n+1}}{T_{n+1} - t_n} \ldots \ldots [19]$$

From Equations [17] and [19] it may be shown that

$$R - \frac{R-1}{\phi} = \frac{T_1 - t_1}{T_2 - t_2} = \frac{T_i - t_i}{T_{i+1} - t_{i+1}}$$

where

$$\phi = \frac{T_{n+1} - t_{n+1}}{T_{n+1} - t_n}$$

Therefore, the ratio of the terminal temperature differences for n tanks is

$$\frac{T_1 - t_1}{T_{n+1} - t_{n+1}} = \left(R - \frac{R-1}{\phi}\right)^n = Z^n \ldots \ldots [20]$$

By adding Equation [18] for every tank and noting the relationship given in Equation [19]

$$wc(t_1 - t_{n+1}) = UA\, \frac{t_1 - t_{n+1}}{\log_e \phi}$$

but $wc(t_1 - t_{n+1}) = UAn\Delta t_m$ by definition, therefore

$$\Delta t_m = \frac{t_1 - t_{n+1}}{n \log_e \phi} \ldots \ldots \ldots [21]$$

which may be expressed in terms of the terminal temperatures.

If R, ϕ, and the inlet temperatures are known, Equations [20] and [17] may be used in calculating the intermediate temperatures.

From Equation [20]

$$T_1 - t_1 = Z^n T_{n+1} - Z^n t_{n+1} \ldots \ldots [22]$$

and from Equation [17]

$$T_1 - T_{n+1} = Rt_1 - Rt_{n+1} \ldots \ldots [23]$$

solve for t_1 in Equation [22], substitute in Equation [23], and simplify to obtain

$$\frac{t_1 - t_{n+1}}{T_1 - t_{n+1}} = \frac{Z^n - 1}{RZ^n - 1} \ldots \ldots [24]$$

These relations can be constructed into a simple graph. In Fig. 14, the line OG is drawn through point T_1, t_1 with a slope of R. The line OH is drawn to divide any horizontal segment GI between OG and the diagonal OI in such proportion that $GI/HI = \phi$. By similar triangles $GI/HI = GE/HJ$, which latter expression is ϕ by definition. The temperature in any tank is found by proceeding stepwise from T_1 and t_1 for n steps.

Tanks in Series; Coils in Parallel. In this case, let t_1 = inlet temperature of all coils and $t_2, t_3, \ldots t_{n+1}$ = outlet temperature for first, $\ldots . n$ tanks.

For any tank, say tank No. 1

$$wc(t_2 - t_1) = UA\, \frac{(T_2 - t_1) - (T_2 - t_2)}{\log_e \frac{T_2 - t_1}{T_2 - t_2}} \ldots \ldots [25]$$

If UA is the same for each tank

$$\frac{UA}{wc} = \log_e \frac{T_{n+1} - t_1}{T_{n+1} - t_{n+1}} \ldots \ldots \ldots [26]$$

From Equations [25] and [26], it is readily found that

$$R\left(1 - \frac{1}{\psi}\right) + 1 = Y = \frac{T_1 - t_1}{T_2 - t_1} = \frac{T_2 - t_1}{T_3 - t_1}, \text{ etc.} \ldots [27]$$
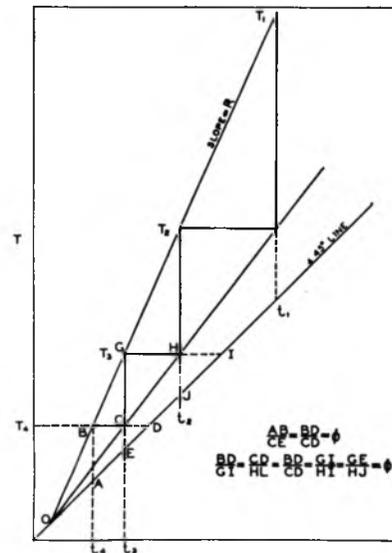
Fig. 14 Diagram for Graphical Solution for Temperatures in Series of Tanks Heated or Cooled by a Series of Countercurrent Coils

where

$$\psi = \frac{T_{n+1} - t_1}{T_{n+1} - t_{n+1}}$$

whence

$$Y^n = \frac{T_1 - t_1}{T_{n+1} - t_1} \ldots \ldots \ldots [28]$$

Batch Heating or Cooling

Many industrial processes are operated batchwise; and where the material is heated or cooled the problem of determining a mean temperature difference varying with time is encountered. Several methods have been derived for calculating heat transfer or the mean temperature difference. Perry (12) presents one of the final equations and attributes the derivation to Boelter and Cherry. The equations were integrated with the following assumptions:

1 The rate of flow and inlet temperature of the heating or cooling medium are constant

2 The specific heat of each fluid is constant.

3 Only sensible heat is involved, there being no change of phase or heats of reaction.

4 The batch fluid is well agitated and of uniform temperature T.

5 Heat losses are negligible.

6 The over-all heat-transfer coefficient U is constant.

The general design equation is

$$Q = UA\Delta t_m \Theta = MC(T_1 - T_2) \quad \ldots \ldots \ldots \ldots [29]$$

and expressed differentially as

$$\frac{dT}{(T-t)_m} = \frac{-UA d\Theta}{MC} \quad \ldots \ldots \ldots \ldots \ldots [30]$$

where $(T - t)_m$ is the log mean of $(T - t_a)$ and $(T - t_b)$, t_a and t_b being the inlet and outlet temperature of the flowing fluid at time Θ. A heat balance on the batch and cooling media at time Θ is

$$wc(t_b - t_a)d\Theta = -MCdT \quad \ldots \ldots \ldots \ldots [31]$$

Solve Equation [31] for t_b and substitute in the log-mean temperature difference $(T - t)_m$ in Equation [30]. This gives

$$\log_e \frac{T - t_a}{T - t_a + \dfrac{MCdT}{wcd\Theta}} = \log_e \frac{\Delta t_a}{\Delta t_b} = \frac{AU}{wc} \quad \ldots \ldots [32]$$

which may be rewritten and integrated over time period Θ, to give

$$Q = wc\Theta(T - t_a)_m \left(1 - e^{\frac{-AU}{wc}}\right) \quad \ldots \ldots \ldots [33]$$

or

$$-UA = 2.3 \, wc \, \log_{10}\left(1 - \frac{Q}{wc\Theta \, (T - t_a)_m}\right)$$

where $(T - t_a)_m$ is the log-mean difference of the terminal batch temperatures and the inlet temperature t_a of the medium.

Either Equation [32] or [33] may be used depending upon the information available. Equation [32] is useful when the temperature rise of the heating or cooling medium is known at any time Θ and batch temperature T. With this equation experimental data may easily be investigated to determine the variation of the over-all coefficient U. If the over-all coefficient varies as a straight-line function of temperature T, then Colburn (13) has shown that the value of $U\Delta t_m$ should be the log mean of $U_1\Delta t_2$ and $U_2\Delta t_1$. Equation [33] is useful in design because it is not necessary to know the exit temperature of the flowing fluid. The foregoing equations were derived on the assumption of complete mixing of the batch at any time; this, however, is seldom realized in actual operation. In the other extreme of no mixing the problem is that of unsteady-state conduction. Methods of determining mean temperature for various degrees of mixing have not been developed.

ACKNOWLEDGMENT

The authors are grateful for permission to include the previously unpublished results of A. P. Colburn for the batch heating process, of T. B. Drew for horizontal film-type exchangers and tanks in series, and of E. F. Yendall on the one-two and one-four heat exchangers.

BIBLIOGRAPHY

1 "Mean Temperature Differences in Multipass Heat Exchangers," by W. M. Nagle, *Industrial and Engineering Chemistry*, vol. 25, 1933, pp. 604–609.

2 "The Calculation of the Mean Temperature Difference in Multipass Heat Exchangers," by A. J. V. Underwood, *Journal*, Institution of Petroleum Technologists, vol. 20, 1934, pp. 145–158.

"Graphical Computation of Logarithmic-Mean Temperature Difference," by A. J. V. Underwood, *Industrial Chemist*, vol. 9, 1933, pp. 167–170.

3 "Mean Temperature Difference Correction in Multipass Exchangers," by R. A. Bowman, *Industrial and Engineering Chemistry*, vol. 28, 1936, pp. 541–544.

4 "Mean Temperature Difference Correction in Multipass Exchangers," by F. K. Fischer, *Industrial and Engineering Chemistry*, vol. 30, 1938, pp. 377–383.

5 "Multipass Exchanger Calculations," by H. Ten Broeck, *Industrial and Engineering Chemistry*, vol. 30, 1938, pp. 1041–1042.

6 "Der Wärmeübergang im Kreuzstrom," by W. Nusselt, *Zeitschrift des Vereines deutscher Ingenieur*, vol. 55, 1911, pp. 2021–2024.

7 "Eine neue Formel für den Wärmedurchgang im Kreuzstrom," by W. Nusselt, *Technische Mechanik und Thermodynamik*, vol. 1, 1930, pp. 417–422.

8 "Mean Temperature Difference in Cross Flow," by D. M. Smith, *Engineering*, vol. 138, 1934, pp. 479–481, 606–607.

9 "Nouvelles Études sur la Chaleur," by Ch. Roszak and M. Véron, Dunod, Paris, 1929, pp. 258–267, 731–750.

10 "Theory of the Single-Pass Cross-Flow Heat Interchanger," by A. M. Binnie and E. G. C. Poole, Proceedings, Cambridge Philosophical Society, vol. 33, 1937, pp. 403–411.

11 "Note sur les Échangeurs à Circulation Croisée," by André Liebaut, *Chaleur et Industrie*, vol. 16, 1935, pp. 286–290.

12 "Heat Transfer in a Horizontal Coil Vat Pasteurizer," by R. L. Perry. Published by THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS in "Heat Transfer," 1935, pp. 61–68.

13 "Mean Temperature Difference and Heat Transfer Coefficient in Liquid Heat Exchangers," by A. P. Colburn, *Industrial and Engineering Chemistry*, vol. 25, 1933, pp. 873–877.

14 "The Calculation of Heat Transfer in Cross-Flow Exchangers," by E. A. Schumann, Jr., contributed to the A.S.M.E. Heat Transfer Professional Group and presented at the Annual Meeting, Philadelphia, Pa., December 4–8, 1939, of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS.

15 Private communication.

Discussion

E. A. SCHUMANN, JR.[5] The writer is well aware of the enormous labor expended in the preparation of the various correction-factor plots given in this paper and feels that the authors are to be thanked for their contribution to heat-transfer-design data.

A comment seems appropriate on the significance of the dimensionless correction factor F in studies of the relative merits of counterflow and unmixed crossflow heat exchangers. It is readily seen that the correction factor represents a ratio of mean temperature differences.

$$F = \frac{\Delta t_m \text{ (crossflow)}}{\Delta t_m' \text{ (counterflow)}}$$

and further, that

$$F = \frac{U'}{U} \times \frac{A'\text{(counterflow)}}{A\text{(crossflow)}}$$

Since F is less than unity

$$UA > U'A'$$

and for identical rates of heat absorption in both types of exchanger, it is apparent that the extent of crossflow heating surface need not be greater than that in the counterflow exchanger, as is sometimes supposed.

On the contrary, designs which can be adapted to high mass-flow rates of the shell-side fluid (tubular-type unmixed crossflow exchanger) commensurate with economical pressure drop may be made to operate with higher shell-side fluid conductances than could be obtained in a counterflow installation.

Finally, any correction factor based on the authors' argu-

[5] Engineer, Construction Department, The Detroit Edison Company, Detroit, Mich. Jun. A.S.M.E.

ment P, and fluid ratio R, is only as reliable as the assumption made regarding the strictness with which the expected or actual flow directions adhere to the ideal crossflow.

Authors' Closure

The intent of this paper is to present a simplified method for determining the true mean temperature difference for various types of flows encountered in heat exchangers from the log-mean temperature difference by applying a correction factor. The correction factor F should not be interpreted as indicating the relative merits of countercurrent exchangers with other flow-type exchangers as this would hold only under the special conditions of identical flow rates, temperatures, and heat-transfer coefficients. While it may be possible by using an exchanger other than a countercurrent type to increase the over-all heat-transfer coefficient more than the lowering of the true mean temperature difference, there are, however, some types of exchangers for which certain temperature conditions cannot be met irrespective of the value of the heat-transfer coefficient; this is shown in example 2 of Table 1. No method of calculating the true mean temperature difference for the crossflow case has been developed for cases where the flow is not either completely mixed or unmixed; this was specified in the assumptions on which the equations are based.

Automatic Control in the Presence of Process Lags

By C. E. MASON,[1] FOXBORO, MASS., AND G. A. PHILBRICK,[2] FOXBORO, MASS.

In this paper, the methods of "Quantitative Analysis of Process Lags" developed in a previous paper[3] are extended to include automatic control. Carrying on the idea of representation by liquid-level systems, a specific system is postulated as a model for a typical industrial process which may or may not be thermal. Continuing the quantitative approach, the mathematical machinery is demonstrated by which the operation of control on processes may be investigated. A selected succession of controllers, characterized by the laws of control which govern them, are considered in connection with their performance on the model system. Although no attempt has been made to formulate rules governing the proper application of the various types of controllers considered or governing the adjustment of any of these individual types, the authors consider that analyses of this sort may well serve as a guide to practical theorems in a technology of automatic control.

THE quantitative behavior of certain resistance-capacity systems was discussed in a former paper[3] presented before this Society. These systems exhibited characteristics similar to those of industrial processes in which control-resisting lags are found, and were thus proposed as convenient models for such processes. To bear out the implied analogy, the models themselves, considered as liquid-level systems, were developed side by side with equivalent thermal systems. In this way it was found possible to give definite analytic form to the all-important dynamic properties of equipment to which the application of automatic control is contemplated, and to justify a professed effort to fill the gap in a literature otherwise primarily concerned with the properties of the controls themselves.

Having already dealt with the inherent attributes of a few such model systems, the authors will show in the present paper, how the treatment may be extended to cover their behavior under automatic control. The procedure will be to choose a specific model process to which to apply a succession of types of controls (assumed to be pneumatically operated), the model being one which is representative of a typically difficult prototype system, involving a combination of capacity and transfer lags. For concreteness, this prototype system may be considered as a thermal process, although it could equally well involve hydraulic, pneumatic, or electrical phenomena, either singly or in combination.

Although the alleged purpose of an automatic-control installation is that of holding constant a certain process variable, this is in reality only accomplished by a series of recoveries from

threatening unbalances, the continuity of which constitutes the record of control. In practice, the disturbances are normally gradual to some degree. It is this fact which makes possible the apparent elimination of the effects of upsetting influences.

Since it is obviously impractical to draw conclusions as to the over-all (transient) characteristics of the combined system through the application of any series of influences, it is desirable to select an isolated, suddenly applied, disturbing influence. The results of such a disturbance must involve the complete transient characteristics of the process and controller in combination.

The model system selected as representative is shown in Fig. 1. The capacity and resistance elements of which it is constructed are the same sort as previously described.[3] T_a, T_b, and T_c are
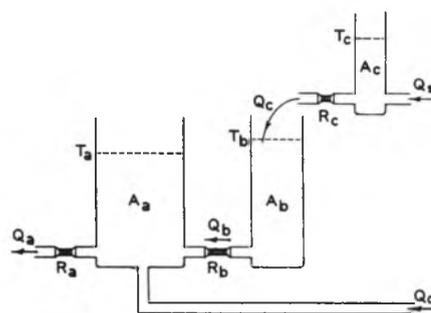
FIG. 1 HYDRAULIC MODEL OF A TYPICAL CONTINUOUS PROCESS

levels (potentials) in the three tanks, the equivalent areas[4] (capacities) of which are A_a, A_b, and A_c. These areas are interconnected through the restrictions (resistances) R_a, R_b, and R_c. The flows Q_a, Q_b, Q_c, Q_s, and Q_o are disposed as shown.

The level T_a in A_a is intended to be the controlled variable, while the supply flow Q_s is to be given the role of manipulated variable by means of which the control of T_a is accomplished. The auxiliary-supply flow Q_o provides a means for altering the load condition, for on the value of this flow depends the value of

TABLE 1 SUMMARY OF PHYSICAL UNITS FOR HYDRAULIC MODEL AND THERMAL PROCESS

	Dimensional symbol	Units	
		Hydraulic counterpart	Thermal counterpart
Quantity	[W]	Pound (weight)	Btu (heat)
Potential (T)	[P]	Inch (level or head)	Deg F (temp)
Time (t)	[T]	Minute	Minute
Flow (Q)	[WT⁻¹]	Lb/min	Btu/min
Capacity (A)	[WP⁻¹]	Lb/in	Btu/deg
Resistance (R)	[W⁻¹PT]	In./(lb/min)	Deg/(Btu/min)

Q_s necessary to produce any given eventual (or potential) value of the controlled variable T_a. The output flow Q_a is directly dependent at any time on the value of T_a and in equilibrium is equal to the sum of Q_s and Q_o.

The physical units for the various constant and variable factors or quantities will be taken as those employed in the former paper,[3] although of course any consistent system of units could

---

[4] A factor proportional to the area, and numerically dependent upon the units used.

be chosen. Table 1 summarizes these units both for the hydraulic model and for the corresponding thermal process.

Since the model shown in Fig. 1 is similar in form to that given as case VI (Fig. 11) in the former paper,[3] differing therefrom only in the inclusion of the cascaded capacity A_c, it will be appropriate to assign the same numerical values to the constants of the system, in so far as this is possible (Table 2). The units to be assumed for these system constants are as given in Table 1, whether for the hydraulic model or for the thermal prototype.

TABLE 2   NUMERICAL VALUES OF SYSTEM CONSTANTS

$A_a = 32.00$	$R_a = 0.1250$
$A_b = 16.00$	$R_b = 0.2500$
$A_c = 4.00$	$R_c = 0.0625$

As pointed out, the original process, of which the hydraulic system of Fig. 1 is to be considered a model, might have any one of a variety of forms. Its nature has been taken arbitrarily as thermal. It might, for example, be thought of as a process identical with the two-capacity thermal system shown in Fig. 4 of the previous paper,[3] but with the modification that the electric heating agency is indirectly manipulated by means of an operation involving simple capacity lag, such as exists with a large air motor setting a current regulator.[5] In any event, for present purposes it may be assumed that the model has been faithfully constructed, and thus the hydraulic system as such may be dealt with directly. Incidentally, the possibility should not be overlooked that the original, or prototype, system might be of the same nature as the model, as for instance a large-scale industrial liquid-level process having transfer lag.

Before proceeding to the application of automatic controls, let us briefly discuss the properties of the specific hydraulic system (or analog) from the point of view of the former paper.[3] Since this is a three-capacity system, the differential equation describing the dynamic relation between the significant variables (T_a, Q_s, and Q_o) will be of the third order. The derivation of this equation will not be given in its entirety since it follows straightforwardly from the methods of the previous paper. The fundamental hydraulic equations are

$$\left. \begin{array}{ll} T_a' = (Q_b + Q_o - Q_a)/A_a & T_a = R_a Q_a \\ T_b' = (Q_c - Q_b)/A_b & T_b - T_a = R_b Q_b \\ T_c' = (Q_s - Q_c)/A_c & T_c = R_c Q_c \end{array} \right\} \dots [1]$$

Eliminating all variables except the significant ones

$$(A_3)T_a''' + (A_2)T_a'' + (A_1)T_a' + T_a = R_a Q_s + R_a Q_o + R_a(B_1)Q_o' + R_a(B_2)Q_o''$$

where
$$\left. \begin{array}{l} (A_1) = A_a R_a + A_b R_a + A_b R_b + A_c R_c \\ (A_2) = A_a R_a A_b R_b + A_c R_c (A_a R_a + A_b R_a + A_b R_b) \\ (A_3) = A_a R_a A_b R_b A_c R_c \\ (B_1) = A_b R_b + A_c R_c \\ (B_2) = A_b R_b A_c R_c \end{array} \right\} \dots [2]$$

which is the complete process equation for the system of Fig. 1. In the present investigation, the auxiliary flow Q_o will be kept always at steady values except when suddenly changed from one steady value to another in simulation of an instantaneous upset; thus for the time intervals considered

$$Q_o' = Q_o'' \equiv 0$$

and the process equation becomes

$$(A_3)T_a''' + (A_2)T_a'' + (A_1)T_a' + T_a = R_a(Q_s + Q_o) \dots [3]$$

[5] In this manner, the so-called "valving" lags may be included as parts of the process rather than as parts of the controls.

Upon insertion of the numerical values given in Table 2

$$4 T_a''' + 18.5 T_a'' + 10.25 T_a' + T_a = 0.125(Q_s + Q_o) \dots [4]$$

in which, by way of recapitulation,

$$\begin{array}{l} T_a = \text{controlled variable} \\ Q_s = \text{manipulated variable} \\ Q_o = \text{auxiliary supply flow} \end{array}$$

The supply flow Q_s is to be operated on by the controlling apparatus in its effort to maintain T_a at a constant value. The auxiliary flow Q_o will suddenly be changed from one value to another under otherwise steady conditions to provide a disturbance to the controlled system at an arbitrarily initial instant ($t = 0$).

As in the former paper, let the desired or normal value $(T_a)_{\text{norm}}$ of T_a be

$$(T_a)_{\text{norm}} = 80 \text{ (in. or deg)}$$

In the thermal counterpart of Fig. 4,[3] the auxiliary flow Q_o was the flow of heat energy—above 0 F—contained in the mechanical flow of 8 lb per min of water. Assuming again that initially the inlet-water temperature is 60 F, then the disturbance can consist of a sudden decrease in this temperature from 60 F to 40 F. Thus, for the thermal system, the flow Q_o had been initially $8 \times 60 = 480$ Btu/min, while the altered value of Q_o is $8 \times 40 = 320$ Btu/min. Hence for either system

$$\begin{array}{ll} (t < 0) & Q_o = 480 \text{ (lb per min or Btu per min)} \\ (t > 0) & Q_o = 320 \text{ (lb per min or Btu per min)} \end{array}$$

Allow the range in which the variable Q_s may be changed by the controls to be given by

$$(Q_s)_{\text{min}} \leq Q_s \leq (Q_s)_{\text{max}}$$

where
$$\begin{array}{l} (Q_s)_{\text{min}} = 0 \text{ (lb per min or Btu per min)} \\ (Q_s)_{\text{max}} = 640 \text{ (lb per min or Btu per min)} \end{array}$$

If the controlled variable T_a is to be at its desired value $(T_a)_{\text{norm}}$ in the balanced or undisturbed preinitial state, then, since the value of the auxiliary flow Q_o is already determined for that epoch, it will be seen from the steady-state version (all derivatives equal zero) of Equation [3] (or Equation [4]) that the initial balanced value of Q_s must be

$$(Q_s)_{\text{init}} = [(T_a)_{\text{norm}}/R_a] - (Q_o)_{\text{init}} = (80/0.125) - 480 = 160$$

Our system is thus prepared for the application of the controlling apparatus which will now be considered.

APPLICATION OF CONTROL APPARATUS

The complete or interconnected control circuit, comprising both controlled process and controller, is symbolically shown in Fig. 2, in which the process is characterized by the group of its inherent system constants, and where the other various components of the controlled system are shown in their functional relationship to one another.

Fig. 3 gives a somewhat more realistic picture of the combined system. In Fig. 1 the elements of the process under control are shown in greater detail.

The manipulated supply flow Q_s is assumed to be under successful control by means of a flow controller, in which instrument the balanced value of flow, or the flow control point, may be set by the operating means (the pneumatic operating pressure) of the level controller, which is the direct analog of a temperature controller in the thermal counterpart. This instrument contains the measuring means and the control mechanism, while the flow controller acts as a positive controlling means. These assumptions will be followed for all types of control mechanism considered.

Suppose for simplicity that, as the operating air pressure P, Fig. 3, of the control mechanism varies from zero to its maximum value, namely, the supply pressure P_s, the resulting supply flow Q_s maintained by the flow controller varies proportionately from zero to its maximum value $(Q_s)_{max}$. Thus

$$\frac{P}{P_s} = \frac{Q_s}{(Q_s)_{max}} \quad\dots\dots\dots\dots\dots\dots [5]$$

With the operating means of the control mechanism thus directly linked to the manipulated flow, the equation of the control mechanism may be expressed as a functional relationship between the controlled variable T_a and the manipulated variable Q_s, or as

$$Q_s = F_c(T_a) = \text{"controller function" of } T_a \dots\dots [6]$$

This corresponds to the form of the process equation which may be written

$$T_a = F_p(Q_s) = \text{"process function" of } Q_s \dots\dots [7]$$

Equations [6] and [7] are of course symbolic since the functions are in general differential expressions and involve time. They
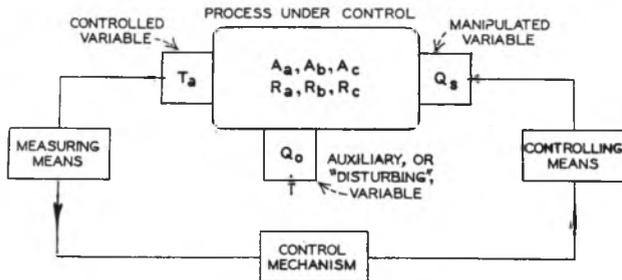

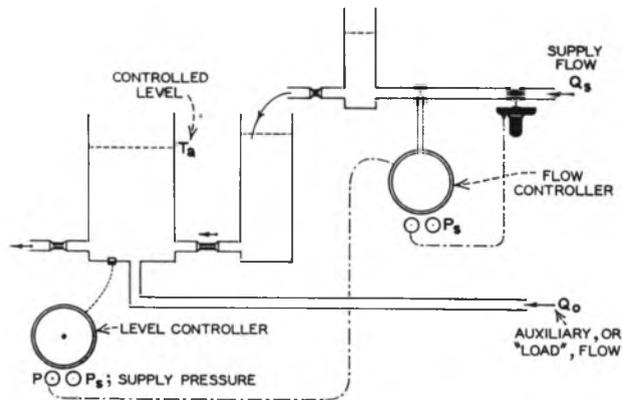FIG. 2   DIAGRAMMATIC CONTROL CIRCUIT

FIG. 3   COMPLETE CONTROLLED SYSTEM

serve, however, to illustrate the reversal of cause and effect in passing from process to controller or from controller to process. In tabular form:

	Cause	Effect
Process	Manipulated variable	Controlled variable
Controller	Controlled variable	Manipulated variable

In the "philosophy" of control, this concept may be considered basic.

The exact nature of the functional relation Equation [6] will depend upon the particular type of controller under consideration. The functional relation Equation [7] for the process is provided in the present case by Equation [3] (and/or [4]).   For

purposes of comparison, not only among the general controller types but between different conditions of adjustment for any one type of controller, the same initial conditions and also the same manner and magnitude of upset will be assumed throughout. Before taking up the controllers in order, it will be informative to determine the effect on the controlled variable T_a of the disturbance to be employed when this disturbance acts on the system in equilibrium and with the controls assumed inoperative. The resulting deviation of T_a will be a standard by which to ascertain the actual amount of correction brought about by the action of the controllers when operative on the process.   Thus, it will be assumed that the foregoing initial conditions are in effect on the process and that, at $t = 0$, the chosen disturbance is inflicted. Since no controller is acting, the supply flow Q_s will remain constant at its initial value, while the auxiliary flow Q_o will have its new balanced value.   Thus from the initial conditions and from the process Equation [4]

$$4\,T_a''' + 18.5\,T_a'' + 10.25\,T_a' + T_a = 0.125\,(160 + 320)$$

or

$$(T_a - 60)''' + 4.625\,(T_a - 60)'' + 2.5625\,(T_a - 60)' + 0.25$$
$$(T_a - 60) = 0 \dots\dots\dots\dots\dots\dots [8]$$

The solution of Equation [8] to find the variation of T_a as a function of time is a procedure similar to those described in the previous paper.[3] It involves the roots of the auxiliary algebraic equation.

$$k^3 + 4.625\,k^2 + 2.5625\,k + 0.25 = 0 \dots\dots\dots\dots [9]$$

and the initial values of the first and second derivatives of T_a (rate of change and acceleration immediately following the disturbance).

The roots of Equation [9] are all real and negative and may be found by any desired method, although in general for speed and precision Graeffe's method[6] is recommended for any linear algebraic equation of higher degree than the second.   As a matter of fact Equation [9] is directly factorable in this case (due to the cascaded nature of the analog process) and may be written

$$(k^2 + 0.625\,k + 0.0625)\,(k + 4) = 0 \dots\dots\dots [9a]$$

The roots are thus

$$k_a = -0.125 \qquad k_b = -0.5 \qquad k_c = -4 \dots\dots [10]$$

The initial values of the first two derivatives of T_a may be determined, through the use of Equations [1], from the change in Q_o constituting the disturbance.   Since the initial value of the third derivative of T_a will also be necessary in later examples and may conveniently be found in conjunction with the other two derivatives, all three have been found for the assumed upset and are listed as follows:

$$
\left.
\begin{aligned}
(T_a')_o &= -5 && \text{(in. per min or deg per min)} \\
(T_a'')_o &= +1.875 && \text{(in. per min}^2 \text{ or deg per min}^2) \\
(T_a''')_o &= -0.859375 && \text{(in. per min}^3 \text{ or deg per min}^3)
\end{aligned}
\right\} \dots [11]
$$

Reference to Appendix 1[7] shows the solution in this case to be of the form

$$T_a - T_p = T_a - 60 = C_a e^{k_a t} + C_b e^{k_b t} + C_c e^{k_c t} \dots\dots [12]$$

[6] "Mathematics of Modern Engineering," by R. E. Doherty and E. G. Keller, John Wiley & Sons, Inc., New York, vol. 1, 1936, pp. 98–130.

[7] The Appendixes referred to throughout the paper form a mathematical supplement which does not appear in this publication. A limited number of copies of this supplement are available, however, and may be obtained from the authors on request.

where the k's are the roots given and the C's are integration constants involving both the roots and the initial values of the derivatives in a manner shown in Appendix 1.[7] Inserting the numerical values there results

$$T_a = 60 + 13.33\,e^{-0.125t} + 6.67\,e^{-0.5t} + 0.00\,e^{-4t}$$

The integration constant C_c is reduced automatically to zero for the case chosen, since the system is essentially two-capacity with respect to changes in Q_o alone, Q_c Fig. 1 remaining constant along with Q_s for the period considered. Thus for the effect of the disturbance on T_a in the absence of control

$$T_a = 60 + 13.33\,e^{-0.125t} + 6.67\,e^{-0.5t} \quad \ldots\ldots\ldots [13]$$

$$\left.\begin{array}{l}\text{Thus when } t = 0 \quad T_a = (T_a)_o = 80 \text{ (in. or deg)} \\ \text{and when } t \to \infty \quad T_a \to (T_a)\infty = 60 \text{ (in. or deg)}\end{array}\right\} \ldots [14]$$

The behavior of T_a following $t = 0$ is shown in Fig. 4. An even-
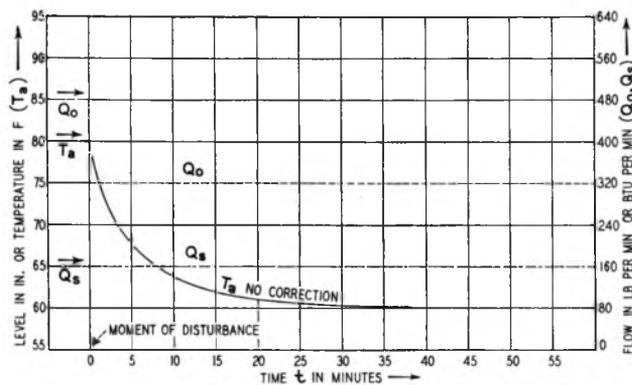

FIG. 4   EFFECT OF DISTURBANCE WITHOUT CONTROL

tual departure of 20 units from the desired value is evident. Plotting may be carried out directly from Equation [13] by means of exponential tables, or through the rather more convenient indirect graphical method used in the previous paper (Appendix 4[7]).

In either manually or automatically controlling this process (Fig. 1), the flow Q_s is changed in an attempt to counteract the effect of the disturbance on T_a, and eventually to restore this variable to its original value. The equation for T_a' (see Equations [1]) and the curve of Fig. 4 show that a change in Q_o has an immediate effect on T_a. On the other hand a change in Q_s, however great, can be shown to have no immediate effect on the value of T_a, due to the intervening parts of the process between the point of application of the manipulated flow Q_s and the location of the controlled variable T_a. Thus no matter what operations are performed on Q_s, a certain inevitable departure of T_a will be experienced, the magnitude of which will depend, not only upon the character (including transfer lags) of the intervening parts of the process, but also on the amount of the available corrective change in Q_s. This leads naturally to the subject of automatic control. First, that type of control will be considered which can produce the greatest initial change in the manipulated variable namely

"TWO-POSITION" CONTROL

Other terms commonly applied to this category are: "on-and-off," "open-and-shut," "fixed-position" control, and the various "....stats." The French *"tout ou rien"* seems particularly picturesque. Certainly there is no need to linger over this characteristic. The law of control may be summarized

briefly, in connection with the system under consideration, as

$$\left.\begin{array}{ll}Q_s = (Q_s)_{\max} = 640 & \text{for } T_a < (T_a)_{\text{norm}} = 80 \\ Q_s = (Q_s)_{\min} = 0 & \text{for } T_a > (T_a)_{\text{norm}} = 80\end{array}\right\} \ldots [15]$$

Since in operation, the value of Q_s is constant except for discontinuous changes involving (theoretically) zero time intervals, the resulting behavior of T_a may be discovered by a series of solutions of the process equation itself using methods similar to those in the foregoing part of this paper and in the previous paper.[3] This procedure involves dividing the time following the disturbance into intervals between successive crossings of the desired value by the controlled variable. Fortunately the linearity of the equations permits superposing additively the individual results of the members of any combination of causes in determining their combined effect. Thus, in the beginning, it is merely necessary to add, to the disturbance deviation given by Equation [13] or Fig. 4, the deviation caused by suddenly increasing Q_s alone from its preinitial balanced value (160) to its maximum value (640). This deviation is similar to curve (a) of Fig. 12 in the former paper and is readily obtained, requiring merely a different set of boundary conditions to be imposed on Equation [4], namely,

$$\begin{array}{ll}Q_s = 640 & (T_a')_o = 0 \\ Q_o = 480 & (T_a'')_o = 0 \\ \quad\quad T_p = 140\end{array}$$

Reference to Equation [1] will show why the first two derivatives are initially unaffected by a sudden change in Q_s. The numerical
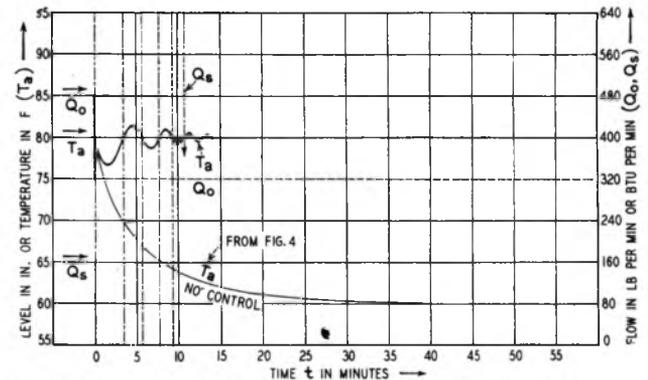
FIG. 5   RECOVERY TRANSIENT WITH IDEAL TWO-POSITION CONTROL

solution, which is found in a similar manner to that for the curve of Fig. 4, is

$$T_a = 140 - 82.59\,e^{-0.125t} + 22.87\,e^{-0.5t} - 0.28\,e^{-4t} \ldots [16]$$

When this deviation is added to that occasioned by the disturbance alone, the first loop of the control transient is obtained, in which T_a passes through its first (inevitable) minimum and returns sharply toward a new potential value. This stage continues only until the desired value is attained, at which time another deviation must be added to the sum of the other two. This is the deviation caused by an independent sudden variation in Q_s from its maximum to its minimum value. This algebraic addition of a new deviation, at each crossing of the desired value (or control-point setting), to the sum of all previous deviations, may be continued as far as is practical or desired. Thus, at approximately 15 min in this case, we are substantially adding together 10 distinct curves. The accuracy of the result becomes increasingly questionable due to the difficulty involved in exact determination of the time intervals. The resulting control transient (or "recovery" transient) is shown in Fig. 5. The

intervals become increasingly shorter as time progresses while the amplitude tends toward zero. Limiting conditions of this sort are only true in an ideal sense, of course, since even slight imperfections, such as dead zones, friction, discontinuity, etc., will make them invalid. It may be noted that, even though the flow Q_s theoretically never becomes steady, its fluctuations are becoming increasingly fast and its average value in this ideal case must approach that required to give T_a the desired value. Thus

$$(Q_s)_{\text{avg}} \rightarrow 320 \text{ (lb per min or Btu per min)}$$

To avoid the irrational assumption that, preceding the disturbance, the flow hypothetically had been in a state of constant motion throughout its entire range, it will be assumed that the process was in equilibrium with the controls inoperative prior to $t = 0$, at which time the controller is made operative simultaneously with the imposition of the upset.

As already noted, the foregoing results apply only to an assumedly ideal two-position control mechanism. This condition may be approached only with a highly sensitive and responsive measuring means. It will be interesting to consider, as an example of what may come under this heading in general, the case of a two-position controller in which a certain degree of dead space is involved, for example in the detection of the controlled variable. Here the control mechanism might not sense the actual value of T_a itself but instead that of another variable in imperfect correspondence therewith. The imperfection meant here is one

FIG. 6 RECOVERY TRANSIENTS WITH TWO-POSITION CONTROLLER HAVING "DEAD SPACE"

which allows the variable corresponding to T_a to follow T_a at a constant separation when the latter variable moves persistently in one direction and to remain unaffected when T_a reverses its direction until T_a has proceeded in the opposite direction by an amount equal to twice the constant separation. This behavior is similar to the action of two imperfectly meshed gears when the rotation of the driver is reversible. Let us determine the effect on the recovery transient.
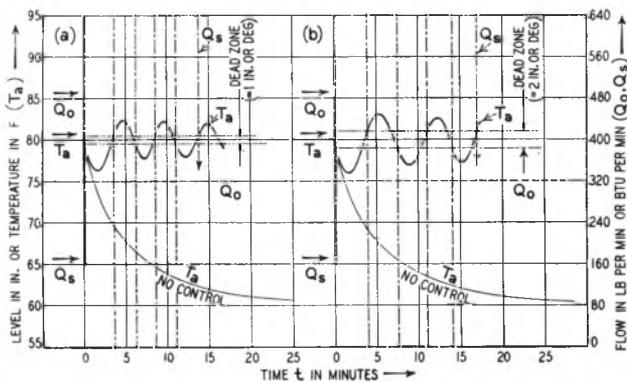
The recovery transient for the two-position controller with dead space may be found in much the same way as for the ideal instrument, except that the correcting changes undergone by Q_s occur, not when the desired value of T_a is crossed, but when T_a has passed beyond that value by an amount equal to $1/2$ the dead zone. This merely requires, in plotting, that the individual effects of the changes involved are added together at different initial instants. Fig. 6 shows the results for 1 and for 2 units (in. or deg) of dead zone, assuming the same initial conditions, etc., as before. The oscillations of T_a can never die out in this case but will eventually settle down to a steady state of oscillation with an amplitude which is always greater than the dead zone and which may even be several times the width of this zone

when complicated process lags are involved. The manipulated variable Q_s will continue to fluctuate regularly between its maximum and minimum values at definite intervals, which may be of objectionable duration.

TWO-POSITION-WITH-RATE CONTROL

This descriptive, though not particularly euphonious, name has been tentatively assigned to a class of control which is widely represented in relatively simple applications. Closely allied to the simple two-position class, it differs therefrom in that it gives to the action of the controlling means a definite constant rate when the controlled variable is on one side of the control-point setting, and gives (usually) the same rate in the opposite
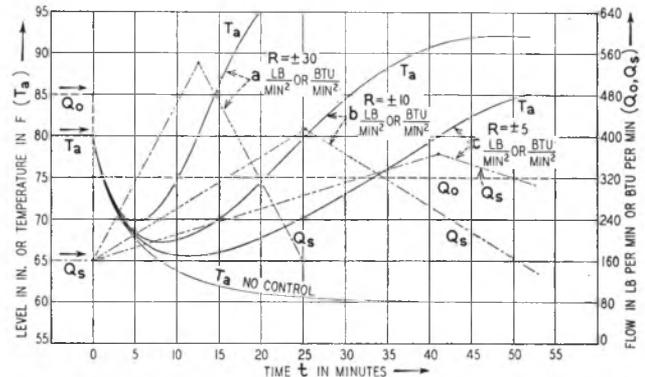
FIG. 7 RECOVERY TRANSIENTS UNDER "TWO-POSITION-WITH-RATE" CONTROL

sense when the controlled variable is on the other side of the control-point setting. This "law of control" may be expressed for the case under consideration, as

$$Q_s' = +R \quad \text{when} \quad T_a < T_{cp}$$

$$Q_s' = -R \quad \text{when} \quad T_a > T_{cp}$$

The quantity R is the constant rate at which the manipulated variable (the flow Q_s) changes. It should be added that this expression of the law requires that Q_s remain between the limits zero and $(Q_s)_{\text{max}}$. This type, it is evident, approaches two-position control as the rate R becomes progressively larger.

This characteristic is included primarily for the sake of completeness; its behavior in the face of process lags is markedly inferior. Thus we shall not give the mathematical procedure by which determination of its behavior on the chosen system may be found, but merely note that this involves solution—by stages—of the nonhomogeneous differential equation

$$(A_3)T_a''' + (A_2)T_a'' + (A_1)T_a' + T_a - R_a(Q_0 + Rt) = 0$$

Results of such determinations are shown in Fig. 7, in which the following rates are assumed:

$$\text{Case } (a) \quad R = 30 \quad \left(\frac{\text{lb}}{\text{min}^2} \text{ or } \frac{\text{Btu}}{\text{min}^2}\right)$$

$$\text{Case } (b) \quad R = 10 \quad \left(\frac{\text{lb}}{\text{min}^2} \text{ or } \frac{\text{Btu}}{\text{min}^2}\right)$$

$$\text{Case } (c) \quad R = 5 \quad \left(\frac{\text{lb}}{\text{min}^2} \text{ or } \frac{\text{Btu}}{\text{min}^2}\right)$$

The relatively fast rate used in case (a) is seen to produce an excessive overshoot on return, while a considerably slower rate, as in case (c), delays recovery from the upset and permits a greater

deviation. It is evident that a compromise, as in case (b), is ineffective in avoiding these two evils.

The inadequacy of this type of control in the present instance is due to the inclusion of transfer lags in the process. This method of control has been used successfully in certain commercial applications, however, but its success has resulted from an abundance of capacity in that part of the process in which the controlled variable is measured. When such conditions prevail, two-position control would be even more successful in the degree of control obtainable, but the two-position-with-rate method allows much smaller fluctuations in the manipulated flow.

ILLUSION OF AN IMMEDIATE, "EXACT" CORRECTION

It may be interesting to note the effect, on the behavior of T_a following an upset, of an enforced immediate change in Q_s to a
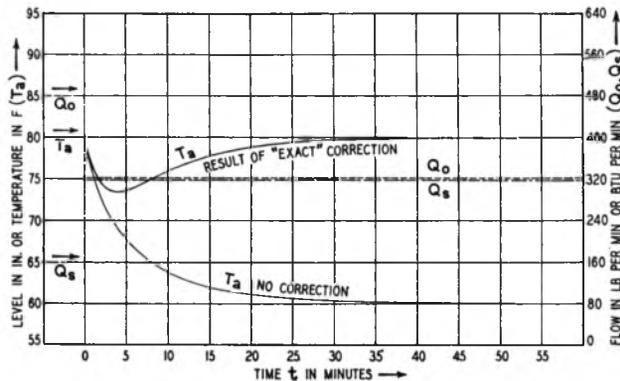
FIG. 8   EFFECT OF AN "EXACT" CORRECTION

new steady value of such magnitude as to produce eventual return by T_a to its desired value. Such a change in Q_s would necessitate a precise knowledge of the disturbance suffered, and might be thought to be an ideal correction. It is ideal, however, only in the steady-state sense. By the foregoing superposition method employed we may plot the resulting deviation by adding to that of Fig. 4 the result of suddenly changing Q_s from 160 to 320 units (an "exact" correction). Fig. 8 shows the consequent behavior of the controlled variable as a relatively slow aperiodic return to the desired value, involving a considerable maximum deviation. This type of correction would be satisfactory, in the transient sense, only if negligible lags existed between the manipulated and controlled variables. Furthermore, it does not lend itself to automatic execution since the measuring means can respond only to the result of a disturbance and cannot measure the cause which provoked it.

PROPORTIONAL CONTROL

This type is also called "throttling" control, due to the definitely allocated band of values (termed the "throttling band") in which the mechanism acts. The operating means of the control mechanism, here the value of the pneumatic operating pressure, is made to correspond linearly with the values of the controlled variable within the throttling band in such a way that the operating pressure has its maximum value when the controlled variable is at the lower limit of the throttling band and has its minimum value when the controlled variable is at the upper limit of that band. Whenever the controlled variable is below (or above) the throttling band, the operating pressure stands at its maximum (or minimum) value. When the controlled variable is precisely in the center of the throttling band, at a value arbitrarily called the "control-point setting," the operating pressure stands halfway between its extreme values.

Since the manipulated variable Q_s is made coextensive with the operating pressure of the control mechanism, the foregoing remarks on the behavior of the operating pressure apply directly to that variable as well. Thus, assuming that the minimum value of the manipulated variable Q_s is zero, the law of control may be expressed as

$$\frac{Q_s - \dfrac{(Q_s)_{max}}{2}}{(Q_s)_{max}} = -\frac{(T_a - T_{cp})}{b}$$

where T_{cp} = control-point setting, or value of T_a at center of throttling band

     b = absolute extent of throttling band in same units as T_a

For completeness the following boundary conditions should also apply

$$Q_s = (Q_s)_{max} \text{ . for } T_a < T_{cp} - (b/2)$$
$$Q_s = 0 \qquad \text{for } T_a > T_{cp} + (b/2)$$

It will be seen that the equations given merely embody the description of the mechanism's operation. Written in a more usable form

$$Q_s = \frac{(Q_s)_{max}}{2} - \frac{(Q_s)_{max}}{b}(T_a - T_{cp})\ldots\ldots\ldots[17]$$

or, with the assumed numerical datum as to $(Q_s)_{max}$

$$Q_s = 640\,[1/2 - (1/b)(T_a - T_{cp})]\ldots\ldots\ldots[18]$$

The control-point setting T_{cp} and the throttling band b are of course "adjustables" of the mechanism. It may be said that the characteristic of this controller is a static one, in that the value of the manipulated variable at any time depends only upon the coincident value of the controlled variable and is independent of all previous or subsequent behavior of the latter variable.

The first step in the investigation of the performance of this type of control in connection with the specific process will be to combine the controller Equation [17] with the process Equation [3], obtaining

$$(A_3)T_a''' + (A_2)T_a'' + (A_1)T_a' + T_a = R_aQ_o + R_a\left[\frac{(Q_s)_{max}}{2}\right.$$
$$\left. - \frac{(Q_s)_{max}}{b}(T_a - T_{cp})\right]$$

or

$$(A_3)T_a''' + (A_2)T_a'' + (A_1)T_a' + (1 + s)(T_a - T_p) = 0$$

where

$$s = \frac{R_a(Q_s)_{max}}{b}\ldots\ldots\ldots\ldots[19]$$

$$T_p = \frac{R_aQ_o + s\,[T_{cp} + (b/2)]}{1 + s}\ldots\ldots\ldots[20]$$

Now since s and T_{cp} will be constant during the time interval considered, the combined equation may be written as

$$(A_3)(T_a - T_p)''' + (A_2)(T_a - T_p)'' + (A_1)(T_a - T_p)'$$
$$+ (1 + s)(T_a - T_p) = 0\ldots[21]$$

The solution of Equation [21] under the usual initial conditions will give the behavior of T_a following the disturbance under proportional control. The meaning of T_p is shown (by Equation

[21]) to be the value at which T_a will eventually balance after the oscillations following the disturbance have faded. The controlled variable will not, except for one particular value of Q_o, balance out at the control-point setting. This general phenomenon is often referred to as "loss of control point." With this type of control the values of $(T_a)_{norm}$, T_p, and T_{cp} are distinct and different except for special cases.

Since the initial value of Q_o is known, Equation [20] will yield the value of T_{cp} necessary to give an initially balanced value of T_a, namely $(T_p)_o$, equal to the desired value $(T_a)_{norm}$. Thus

$$T_{cp} = \frac{(1 + s) (T_a)_{norm} - R_a Q_o}{s} - \frac{b}{2} \dots \dots [22]$$

Two separate cases will be considered in which the throttling bands are

$$(a) \dots b = 10 \text{ (in. or deg)}$$
$$(b) \dots b = 40 \text{ (in. or deg)}$$

Corresponding to these we have, from Equation [19]

$$(a) \dots s = 8 \text{ (dimensionless sensitivity constant)}$$
$$(b) \dots s = 2 \text{ (dimensionless sensitivity constant)}$$

Assigning to $(T_a)_{norm}$ the usual value of 80, Equation [22] gives, for the two cases, respectively

$$(a) \dots T_{cp} = 77.5 \text{ (in. or deg)}$$
$$(b) \dots T_{cp} = 70.0 \text{ (in. or deg)}$$

Thus, from Equation [20], the final balanced values are

$$\left. \begin{array}{l} (a) \dots T_p = 77.78 \text{ (in. or deg)} \\ (b) \dots T_p = 73.33 \text{ (in. or deg)} \end{array} \right\} \dots \dots [23]$$

The combined differential Equation [21] for controller and process becomes, respectively, in each of the two cases

$$4(T_a - T_p)''' + 18.5 (T_a - T_p)'' + 10.25 (T_a - T_p)'$$
$$+ 9 (T_a - T_p) = 0 \dots [24]$$

and $4(T_a - T_p)''' + 18.5 (T_a - T_p)'' + 10.25 (T_a - T_p)'$
$$+ 3 (T_a - T_p) = 0 \dots [25]$$

where the values of T_p are given by Equation [23].

The solutions of Equations [24] and [25] involve cyclically varying functions of time. The process of solution differs from

FIG. 9 RECOVERY TRANSIENTS WITH PROPORTIONAL CONTROL

those cases encountered heretofore in that the roots of the auxiliary algebraic equation are not all real. Appendix 2[7] shows a formal method of integration which may be used here, and which gives, for the two examples

$$T_a = 77.78 + 0.0472 e^{-4.137t} + 6.549 e^{-0.2440t} \cos (0.6959t$$
$$+ 1.231) \dots [26]$$

and $T_a = 73.33 + 0.0120 e^{-4.036t} + 11.602 e^{-0.2944t} \cos (0.3149t$
$$+ 0.9599) \dots [27]$$

These equations describe the behavior of T_a following the instant of upset and are plotted (Appendix 4[7]) in Fig. 9. It is essentially the exponential-cosine (or damped-sinusoid) term of the solution which governs the nature of the oscillations which the control transient comprises. The frequency (or period) and the damping factor are implicitly contained in this term.

For any given process, the roots of the auxiliary algebraic equation, and hence also the nature of the resulting damped sinusoid, are dependent exclusively upon the dimensionless constant s, which is the only variable of the controller entering the coefficients of Equation [21]. The factor s, which has been called the sensitivity constant of the combined system, depends directly upon both the output resistance R_a of the process and the maximum value $(Q_s)_{max}$ of the manipulated supply flow, and is inversely proportional to the extent of the throttling band b, Equation [19]. As the numerical value of the constant s is increased toward a value which would produce discontinuity for the particular upset, the period of cycling becomes shorter, the eventual departure from the desired value $[T_p - (T_a)_{norm}]$ becomes less, while the variation of the manipulated variable Q_s increases. The throttling band b is ordinarily an adjustable of this type of control mechanism, and is used to influence the sensitivity constant s. In practice, an optimum value for s (or for $1/b$) depends upon the permissible variation of Q_s for normal disturbances, and upon the mechanical perfection of the attendant apparatus.

Determination of the fluctuations suffered by the manipulated variable Q_s is fairly direct in this case since this may be accomplished from the already discovered behavior of T_a (Equations [26] and [27]) through the use of the controller Equation [17] or [18]. These fluctuations are also plotted in Fig. 9.

The effect of inertness or dead space in this type of control mechanism would be worthy of very intensive study, due to its inclusion in varying degrees in most commercial instruments. In the case of wide throttling bands, or when the magnitude of dead space is small compared to the throttling band, the effect is principally a proportionate increase in the amplitude and period of oscillation. When the dead space is commensurate with the throttling band, however, the effects may become far-reaching. This condition is especially threatening for small throttling bands; it will be seen that the detrimental effect of dead space as shown in the case of open-and-shut control becomes more influential as the throttling band approaches zero.

Once a substantial balance is attained with proportional control having dead space, the insensitivity of the controlling means to changes of the controlled variable within the dead space may result in complete inactivity of the controls during intervals of minor disturbances. Following larger disturbances, however, intervals of violent cycling can occur which may persist even after the cessation of the upsets. Many inexpensive regulators of this nature are in commercial use which apparently function ideally under small or slow disturbances, but which, when changes occur, calling for significant action by the control mechanism, seem to "lose their heads" and require manual assistance to regain their requisite inactivity.

"FLOATING" CONTROL

The name "integrating control" is sometimes given to this classification for reasons which will be evident. Except for the fact that some of the less complex regulating devices have approximately this characteristic, it is not in common use today. It is important, however, as a component of the more complex types, and has some unique features.
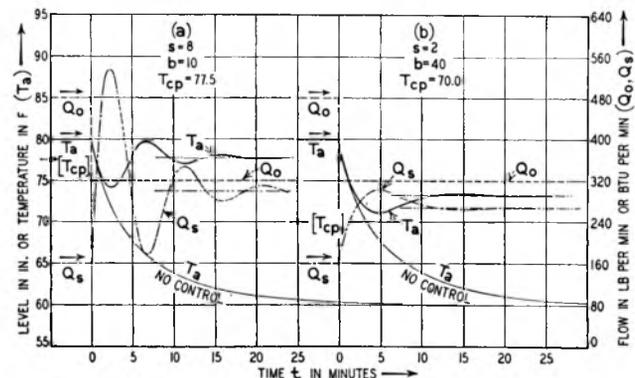
The functioning of this class of mechanism may be described by the statement that the action of the operating means of the control mechanism is a linear function of the integral of the deviation of the controlled variable from a desired value. Taking this value as the control-point setting T_{cp}, we may write, in terms of the operating pressure and using a proportionality factor c

$$\frac{P}{P_s} = -\frac{1}{c} \int (T_a - T_{cp}) dt$$

or, with a definite integral

$$\frac{P}{P_s} - \left(\frac{P}{P_s}\right)_0 = -\frac{1}{c} \int_0^t (T_a - T_{cp}) dt \ldots \ldots [28]$$

A more universally pleasing form is obtained merely by differentiating either of these equations with respect to time; thus

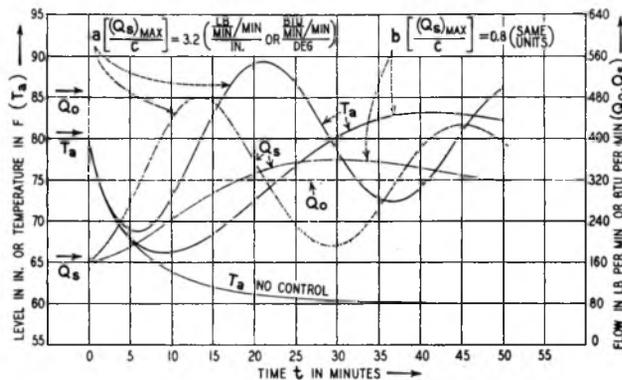$$\frac{P'}{P_s} = -\frac{1}{c} (T_a - T_{cp}) \ldots \ldots \ldots \ldots [29]$$

FIG. 10 RECOVERY TRANSIENTS UNDER FLOATING CONTROL

By the continued assumption that the manipulated variable is made coextensive with the operating pressure, we have from identity [5] and Equation [29]

$$Q_s' = -\frac{(Q_s)_{max}}{c} (T_a - T_{cp}) \ldots \ldots \ldots \ldots [30]$$

Thus the floating characteristic may be described as one in which the rate of change of the manipulated variable is made proportional to the deviation of the controlled variable from a desired value. The constant factor $(Q_s)_{max}/c$ may be thought of as the rate per unit deviation, or as

$$\left[\frac{\frac{lb}{min}/min}{in.}\right] \quad or \quad \left[\frac{\frac{Btu}{min}/min}{deg}\right] \ldots \ldots [30a]$$

The negative sign in Equations [28] to [30] assures the occurrence of the corrective action in the appropriate direction, and allows the proportionality factor to be an inherently positive constant.

The combination of the controller Equation [30] with the time derivative of the process Equation [3] gives the following fourth order differential equation (Q_o being constant disappears in differentiating):

$$(A_3)T_a'''' + (A_2)T_a''' + (A_1)T_a'' + T_a' + q(T_a - T_{cp}) = 0$$

where

$$q = \frac{R_a(Q_s)_{max}}{c} \ldots \ldots \ldots \ldots [31]$$

or, since T_{cp} is constant

$$(A_3)(T_a - T_{cp})'''' + (A_2)(T_a - T_{cp})''' + (A_1)(T_a - T_{cp})'' + (T_a - T_{cp})' + q(T_a - T_{cp}) = 0 \ldots [32]$$

A solution of Equation [32] is possible by methods similar to those used previously and specifically indicated in Appendix 3.[7] Two cases will be considered, in which different values are assigned to c. Before giving the results we may note a fact of great importance from Equation [32]. When the effects of all disturbances are balanced out, all derivatives become zero, and the equation shows that under such conditions $T_a = T_{cp}$. Hence this control characteristic eliminates any permanent or residual effects of changes in load; thus there is no "loss of control point." Although shown here in a specific case, this recognized fact is applicable in general. The shortcomings of this controller type lie in its inability to cope with the immediate effects of a disturbance.

Shown in Fig. 10 are the recovery transients, following the same disturbance as used throughout, for the following adjustments

(a) .... $(Q_s)_{max}/c = 3.2$ (units given by Equation [30a])

(b) .... $(Q_s)_{max}/c = 0.8$ (units given by Equation [30a])

It is evident, from Fig. 10 that, for a larger value than in (a) the extent of the first return loop, or first maximum, would become excessive. Instability, or oscillation of increasing amplitude, lies not far in that direction. For smaller values than in (b), on the other hand, the recovery would involve both an excessive deviation and an excessively slow return. At a certain smaller value of this adjustment, the recovery would become critically aperiodic; that is, the return curve would approach the desired value at the most rapid rate possible without crossing.

Only in the relative absence of transfer lags between the manipulated and controlled variables could the value of $(Q_s)_{max}/c$ be made sufficiently large to give satisfactory sensitivity to the controlling means.

The behavior of the manipulated variable Q_s, as shown in Fig. 10, may be obtained by solution of the original equations for Q_s instead of T_a, or more conveniently (although less accurately) by graphical integration from the recovery curves of T_a (see Equation [28]). The figure plainly shows the lagging phase relationship borne by Q_s to T_a, and that the rate of change of the manipulated variable is zero whenever the deviation is zero.

PROPORTIONAL-PLUS-FLOATING CONTROL

This is a type of control having a characteristic which combines properties of the proportional type and the floating (or integrating) type. Other names are, "throttling-plus-floating," "throttling-plus-reset," "proportional reset," etc. The use of this characteristic is fairly well known today. The success which practical controllers of this type have met with, and which they shall continue to enjoy, may be attributed to the fact that they inherit the particular advantages of their component devices, while the separate or distinct disadvantages of each are fortunately canceled in the combination.

To bear out this conception of its origin, we may show how the equation describing the law of control of this mechanism is synthesized from those of the proportional and of the floating controllers, namely, from Equations [17] and [30]. Differentiating Equation [17]

$$Q_s' = -\frac{(Q_s)_{max}}{b} (T_a - T_{cp})' \ldots \ldots \ldots \ldots [33]$$

Equation [33] states that the rate of change of the manipulated variable is proportional to the rate of deviation of the controlled

TABLE 3 NUMERICAL VALUES OF ROOTS AND INTEGRATION CONSTANTS

Case	r	s	(b)	$+k_a$	$-k_b$	$-k_c$	$-k_d$	$+C_a$	$+C_b$	$-C_c$	$+C_d$
(a)	0.1335	13/12	73.85	0.07791	0.2210	0.1639	4.019	0.5822	54.95	45.91	0.00647
(b)	0.1335	2	40	0.2661	0.2255	0.1390	4.035	1.158	18.09	7.268	0.01165
(c)	0.1335	4	20	0.4448	0.2014	0.1356	4.069	1.317	10.93	2.763	0.02212
(d)	0.1335	8	10	0.6700	0.1788	0.1344	4.133	1.422	8.089	1.239	0.04013
(e)	0.4452	4	20	0.4831	0.04836	0.4648	4.063	−1.253	− 8.362	−2.594	0.02056
(f)	0.4452	1/2	160	0.1559	0.6362	0.4896	4.008	−1.225	−17.15	−5.805	0.00285

variable; this is merely a paraphrasation of the characteristic of the proportional controller. Now if Equations [30] and [33] were in effect additively, we should have, for the new value of Q_s'

$$Q_s' = -\frac{(Q_s)_{\max}}{b}(T_a - T_{cp})' - \frac{(Q_s)_{\max}}{c}(T_a - T_{cp}) \ldots [34]$$

Equation [34] is the law of control for the proportional-plus-floating mechanism.[8] It shows the rate of change of the manipulated variable depending additively upon the deviation of the controlled variable from the control-point setting (desired value) and on the rate of change of that deviation. This equation may also be written

$$Q_s' = -\frac{(Q_s)_{\max}}{b}[(T_a - T_{cp})' + r(T_a - T_{cp})] \ldots [35]$$

in which

$$r = b/c \ldots [36]$$

The constant r may be called the "reset constant." From Equation [35] it is apparent that this term governs the ratio between the influence of the deviation and that of the rate of deviation, or the ratio of the proportional to the integrating effect. The integrating (or "floating") influence brings about eventual equilibrium of the controlled variable at the desired value, and this action has been called "reset." The dimensions of r are (1/time), and its units are "inverse minutes." The constant b is the familiar throttling band of the proportional controller and has, when referred to immediate changes, the same meaning in the present mechanism. Both b and r (and of course T_{cp}) will be considered as "adjustables" of the mechanism.

To obtain the differential equation of the interconnected system, we combine Equation [3] for the process with Equation [35] for the controller, in a similar manner to that followed in obtaining Equation [32] for the integrating case, and obtain thus

$$(A_3)(T_a - T_{cp})'''' + (A_2)(T_a - T_{cp})''' + (A_1)(T_a - T_{cp})'' + (1 + s)(T_a - T_{cp})' + (rs)(T_a - T_{cp}) = 0 \ldots [37]$$

where s is given by Equation [19] and depends inversely on the throttling band, while the role of r has been just described. This equation describes the behavior of the controlled variable T_a under any circumstances (within limits of continuity) and may be solved under the usual specific conditions to show recovery transients obtained with this type of control. Several solutions of this sort will be carried out for various values of the adjustable constants r and s (or of $1/b$). It may be shown that there are rather definite values for r and s to give critically aperiodic return. Strictly definite values giving this critical state would be conveniently obtainable if the combined Equation [37] had been of the third order. This equation would have been of the third order but for the inclusion of the valving lag described earlier and represented in the hydraulic analog, Fig. 1, by the cascade capacity. Hence, by neglecting this lag, values may be computed for the constants r and s which are approximately critical for the complete system. Adjustments of r and s (or of $1/b$), below those values which would give critically aperiodic return, produce a control characteristic which has been called "averaging control."

The effect obtained is to reduce the fluctuation of the manipulated flow at the expense of variations in the controlled variable. It has found wide practical application in level-control problems. Except in such applications, however, the adjustments giving aperiodic behavior following a disturbance are not the most favorable.

In the periodic, or usual, cases Appendix 3[7] shows the solution of Equation [37] to be

$$T_a = T_{cp} + C_d e^{k_d t} + C_c e^{k_c t} + C_b e^{k_b t} \cos(k_a t + C_a) \ldots [38]$$

where the C's and the k's depend upon the coefficients of Equation [37] and hence are affected by the reset constant r and by the sensitivity constant s (or the reciprocal of the throttling band). Table 3 shows the numerical values of the roots and integration constants, for substitution in Equation [38], which result when certain chosen values are assigned to the constants r and s. Case (a) involves the values of r and s which give the substantially critical case, in which neither the magnitude of the throttling band nor the reset constant are such as to cause cyclic behavior. The recovery transient for case (a) is plotted as curve (a) in Fig. 11. Except for a greater deviation, curve (a) exhibits a similar recovery to that resulting from an exact instantaneous correction as described previously and plotted in Fig. 8. This similarity is due to the fact that the flow attains its final value essentially without overshoot in the comparatively brief interval of 11 min. Curve (a) of Fig. 12 shows the recovery transient
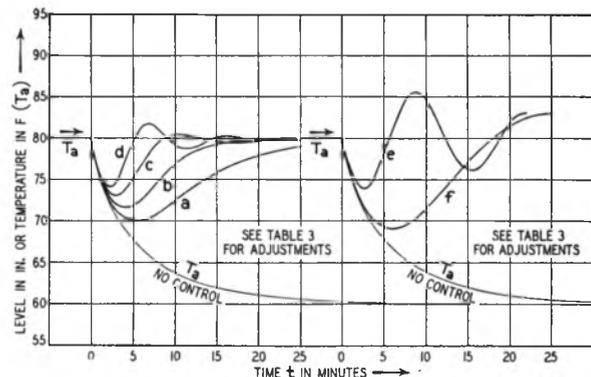

FIG. 11 RECOVERY TRANSIENTS UNDER PROPORTIONAL-PLUS-FLOATING CONTROL; CONTROLLED VARIABLE

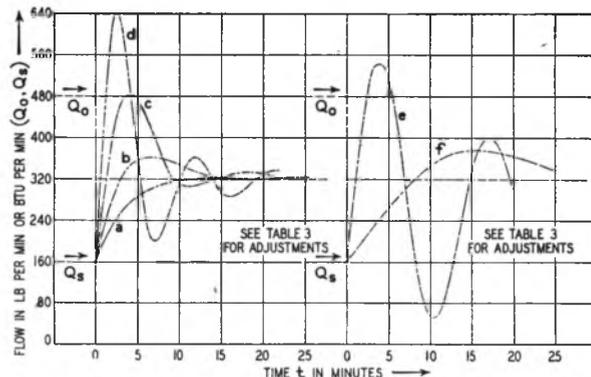FIG. 12 RECOVERY TRANSIENTS UNDER PROPORTIONAL-PLUS-FLOATING CONTROL; MANIPULATED VARIABLE

[8] This equation could also have been developed from the mechanics of a particular control device faithfully embodying proportional-plus-floating characteristics.

for the manipulated flow in this case. The flow transients for this type of control may be found from the transients of the controlled variable by methods indicated previously.

The remaining cases selected for illustration were not chosen at random, for the results of such a procedure might well be confusing. The adjustments, giving the recovery curves shown, form an orderly sequence which, it is felt, will serve to demonstrate their separate influence on the control behavior.

Assuming for the moment that the adjustments giving the critical case (curves (a), Figs. 11 and 12) are in effect, let us consider, in a qualitative sense, the result of allowing b to remain at its critical value and of changing the adjustment r. Without resorting directly to actual solution, we may observe that when under such conditions, r is made smaller (slower reset rate), a longer time will be required for the manipulated flow to attain a substantially balanced state close to its final, or potential, value (320 lb/min or Btu/min). It is apparent, from the discussion previously given of the critical values of b and r, that a periodic recovery transient will result if the constant b is decreased, or the constant r increased, from its critical value. Considering separately an increase in the reset constant r, it is evident that such an increase will bring about an earlier crossing, by the manipulated flow Q_s, of its final or potential value. These relationships are best discernible in the light of the individual characteristics of the component devices, namely, of the floating and the proportional mechanisms, as exhibited in Figs. 10 and 9, respectively. The control influence of the floating part of the characteristic tends to make the rate of change of the manipulated flow proportional to the deviation of the controlled variable from its desired value; this influence tends to prevent reversal of that rate until the controlled variable has crossed its desired (and potential) value (Fig. 10). On the other hand, the control influence of the proportional part of the characteristic tends to reverse the rate of change of the manipulated flow in coincidence with the reversal of the controlled variable (Fig. 9). The desirability of this latter effect suggests that it might be informative to show the effect, on the recovery transients, of progressive reduction in the magnitude of the throttling band b (or of progressive increments in the sensitivity constant s) with the reset constant r remaining at its critical value.

The results of these adjustments are plotted in Fig. 11. Given in Fig. 12 are the curves showing the behavior of the manipulated variable in the corresponding cases. Starting with the practically aperiodic case (a), the reset constant r is left at its critical value and the sensitivity constant s is increased in geometrical steps above its original value. This initial series of adjustments may of course be considered to be obtained by successive reductions in the throttling band b. As s is increased, or as the throttling band is decreased, the cyclic period and the damping per cycle both become smaller, for the oscillations of Q_s as well as of T_a. The initial deviation of T_a is reduced by the successive adjustments, but the corresponding fluctuations of Q_s are seen to increase in magnitude. In each of the first four cases the promptness of return to the region of the desired value is creditable.

With this type of control, discontinuity should be avoided; thus, the choice of an optimum value, for throttling band (b) for any given process, will also be influenced by the available range of the manipulated variable Q_s as well as by the desired stability of that variable and the mechanical perfection of the apparatus.

Smaller values of r, or reset rates slower than the critical, would merely delay the return in each case. The effect of larger values of r (faster reset rates) is indicated by case (e), in which the magnitude of the throttling band is left the same (20 units) as in case (c), but in which the reset constant is increased approximately threefold. Although this adjustment slightly decreases the period of oscillation, as well as the magnitude of the initial

deviation, the amplitude and damping of the subsequent oscillation is seriously impaired. Herein lies the only real danger arising from haphazard manipulation of this adjustment in the commercial use of this type of control mechanism.

From field experience derived from the application of control mechanisms which involve a proportional control effect (or which have a throttling band, in the usual sense), a general rule has evolved to the effect that cycling can be reduced or eliminated by the expedient of an increase in the magnitude of the throttling band. In the case of the proportional-plus-reset control mechanism, this rule is only of practical value provided that the reset rate (or the reset constant r) is reasonably near the critical value demanded by the lag characteristics of the process under control. Case (f) shows the futility of increasing the extent of the throttling band in an attempt to attain satisfactory control when the reset rate is excessively fast for the process under consideration.

The authors' experiences in the mathematical investigation of other types of processes, as well as actual field experience, indicate that there exists—for every process—a critical reset rate (or reset constant r) which is independent of the disturbances to which the process may be subjected and of the throttling band which may be employed in the mechanism.

In spite of the necessarily limited generality of the mathematical approach in the present paper, in that a single specific process has been dealt with, the authors suggest that it may at least serve to introduce a method by which the development of rational theorems may be made possible on an analytic basis. It is evident that the formulation of rules sufficiently practical for industrial use must be the result of sincere cooperative effort on the part of the technical groups both of the users and of the manufacturers of industrial control equipment.

For the present, experience alone can supply the necessary criteria in the application and adjustment of the more complex control mechanisms. Even this brief introductory treatment, however, would seem to establish certain of the properties characteristic of the reset adjustment. It is evident that, although an optimum reset rate may exist for any process, the determination of this optimum is not a simple routine deduction; furthermore, the usefulness of the reset feature in eliminating "loss of control point" is seen to be limited by the critical nature of its adjustment at values below the optimum.

This paper was prepared, as a continuation of the previous paper,[3] to give, together with that paper, a more complete exposition of a suggested method for a quantitative approach to the development of a science of automatic control. The authors wish to submit, for consideration by the A.S.M.E. Committee on Industrial Instruments and Regulators, the suggestions embodied in both papers, and also to offer their sincere cooperation in the accomplishment of the purposes of the committee. In this connection, they will equally welcome comment or critical discussion from any individuals interested in these purposes.

Discussion

E. S. SMITH[9] AND C. O. FAIRCHILD.[10] This paper presents a valuable and straightforward treatment of a combined controller and plant, i.e., a controlled system for several familiar types of controllers and a relatively uncommon process or plant. While level control is used as a complete analog of temperature control, this does not appear to be adequate since the latter involves the important factor of attenuation due to "distributed capacity and resistance."

Since a process is a method, not an apparatus, the word "plant" should be used instead of "process" when the apparatus for carrying out the process is meant. The use of a volume unit instead of a weight unit as the basic "quantity" unit would seem preferable as appreciably simplifying the dimensional treatment.

The so-called philosophy of control cited in connection with Equation [7] does not appear to be entirely consistent with the essential unity of the plant and its controller implied in footnote,[5] the noted reversal being simply a matter of opposite sign or 180-deg phase difference.

Under two-position control, the conclusions that the intervals and amplitudes progressively decrease with time do not appear to be of value for the common practical case where metering lag exists, in which case the intervals and amplitudes become constant.

The two-position-with-rate control is more commonly known as constant-speed floating control. In Fig. 7 and its context, it is unfortunate that R should be used in another sense than resistance as it was elsewhere used in the paper. The last paragraph of the "Two-Position-With-Rate" section appears to be incorrect or inadequate, since a dead zone or meter friction is necessary to eliminate hunting, and an increase of capacity requires a slowing of the rate R of valve movement to avoid reaching the limits of valve travel.

Under the heading "Proportional Control," the s of Equation [19] might well be explained further as being the ratio of the head T_a for $Q_a = (Q_s)_{max}$ to the head of the throttling band b. Following Eq. [21], the "loss of the control point" is more commonly known as "drift" or "load error." In the absence of further explanation of T_p, T_{cp}, and $(T_a)_{norm}$ the final values of Equation [23] appear to be in error. The generality of the conclusion, "As the numerical value of the constant s is increased toward a value which would produce discontinuity for the particular upset, the period of cycling becomes shorter ....," must be questioned, since this does not seem to be so in temperature control, including where attenuation is involved. If the flow scale of Figs. 9 and 10 were inverted as in a paper[11] by one of the writers, a clearer picture of lags for proportional and floating control would be obtained. Thus, it is immediately apparent that the final control element moves in phase with the measured value of the head T_a in Fig. 9 and lags 90 deg behind such head in Fig. 10. The last sentence of this section apparently means that the throttling control simply breaks down into two-position control under unfavorable conditions.

Under the heading "Floating Control" (of a particular kind), a fuller treatment of q in Equation [31] in terms of c would give the reader a needed mental picture of this before he reaches Equation [36]; at least the dimensions of q and c should be stated. The use of the inverse of c, i.e., $1/c$ seems to the writers to be more nearly consistent with the speed of reset as used throughout the latter part of the paper. The reference to transfer lags in the

penultimate paragraph of this section might well include a reference to the portion of Fig. 1 which is meant. The last paragraph might likewise state that the lag of a floating valve is 90 deg behind that of a valve with proportional control.

In connection with Equations [34] to [36], it might be worth while to bring out that $1/c$ is simply the speed of reset dimensionlessly expressed in terms of the maximum flow per unit of head departure $T_a - T_{cp}$, and that an increase of r is accompanied by a resultant tendency for the stability to decrease. Exception may be taken on the grounds of overgenerality to the statement, "Except in such applications (averaging level control), however, the adjustments giving aperiodic behavior following a disturbance are not the most favorable," a case cited in an unpublished paper by R. P. Lowe and one of the present writers presented before the Society in 1938 constituting an important exception in which the controller had to follow a storm flow smoothly.

The last paragraph of the first column on page 304 appears to imply that the throttling band b and the reset speed $1/c$ are adjusted together to maintain r, or b/c, constant. In the second paragraph beyond this, a question arises as to whether $1/c$ or r is meant. In the following paragraph, it would seem to be the users' turn next if the progress of the present paper is to be consolidated without the possible loss of anything more than a bit of higher mathematics.

While the addition of a cascaded reservoir A_s enabled the authors to show that their approach was not limited to simple cases, yet it is likely that a somewhat simpler plant would provide a clearer relative picture of the merits of the several sorts of controllers or regulators. This suggestion is borne out by the authors' simplification of the treatment of Figs. 11 and 12.

A somewhat different approach may be more helpful in certain cases. For example, the analysis of Ivanoff for temperature control of plants involving thermal lags and attenuation also led to some fairly general conclusions as to reset speed which have considerable practical justification. For another example, Ivanoff also determined a controlling apparatus to fit a desired law of control for a particular plant which presented unusual difficulties.

Time is of the essence in detecting a change and initiating a movement of a final control valve properly to offset the cause of the change. With a proportional controller, which operates without serious metering lag or dead zone, nonhunting throttling control exists with a single-capacity plant, since any change of head immediately acts to position the valve properly. In level control, it may be noted that the addition of a rate-of-departure component permits the controller to act as though it were governed by the net discharge from the reservoir and hence to set the control valve immediately to the proper position to maintain the set level, after a change in the rate of either inflow or outflow, without waiting for the level itself to depart objectionably. The addition of the rate component implies a high sensitivity which is generally obtainable only by the servo-operation of the portion providing such component, which portion may be actuated by either the meter or the control valve as convenient. This "departure-plus-rate-of-departure" controller thus functions in an entirely different manner from that of a throttling control and tends to be more stable; although theoretically, with a single capacity, the valve of a throttling controller will be stable even though its throttling band be indefinitely reduced. However, it is better in practice from the standpoint of stability to have a reasonably wide "initial throttling band" followed by a gradual reset to the set point.

The addition of one cascaded capacity also requires the addition of a rate component to let the control valve be properly positioned soon after a variable supply has changed, considering that the outflow be controlled.

From the standpoint of promptness of response, the similar addition of a second cascaded capacity seems similarly to require the addition of an acceleration component, and so on. A general rule seems to be that an increase in attenuation in the number of cascaded capacities or in the order of the differential equation for the plant requires the corresponding addition of a higher derivative to the controller to make prompt control with accurate compensation possible. The use of a floating control valve instead of a throttling control valve likewise raises the order and requires the addition of a higher derivative. An increase in the power of time of the law of departure likewise in effect raises the order and necessitates the addition of a higher derivative if control is not to be lost while within the range of the controller. From these particular considerations, it appears that a controller may be straightforwardly selected to give the desired mode of control of a plant instead of trying empirically a number of different sorts of controllers in an effort to find one which has an acceptable performance with the particular plant.

The reset requirements of flow-rate controllers are of a different order from those of temperature controllers. If the authors have space in their closure, they might include a statement as to whether their more general remarks upon the effects of the reset speed on stability apply equally to flow controllers. The effects of cavitation with gassy liquids and of water hammer in long lines may well modify radically some of the general conclusions with flow-rate controllers where a premium may exist on making the flow changes as gradual as possible. This requirement also exists in other cases such as that in which a thermal gradient must be maintained without exceeding a critical temperature.

Since there is no true distinction in a control system between the plant and its controller and the paper includes a comparative treatment of controllers, it seems desirable to call attention briefly to generally related patent literature: Patent No. 2,113,-164 of A. J. Williams, Jr., contains an interesting analysis of a system which involves both the first and the second derivatives; and the C. E. Mason patent No. 1,897,135 (Re. 20,092) is for a device using the first derivative and having a coupling of a differential bellows to its nozzle and flapper. The latter also includes an interesting analysis. It is requested that the author's closure state whether this Mason patent for the "Stabilog" controller should be regarded substantially "proportional-plus-(first derivative) floating control" or as having a significant higher derivative effect due to close coupling, which effect is not included in the mathematical analysis of such patent.

W. D. WOOD.[12] In the section of the paper covering "Proportional Control," the authors have used the term "Control band," and have developed from it a dimensionless sensitivity constant s. This proportional sensitivity constant is familiar to the readers of the published works of Ivanoff, who also derived the term "potential temperature." The sensitivity constant s may be simply defined as the potential temperature change caused by unit temperature change as measured by the control instrument. This constant may be aptly termed "proportional over-all sensitivity," since it includes the sensitivities of all elements in the control circuit. Ivanoff has pointed out the relation between this sensitivity constant and "loss-of-control point" in proportional control. It is worth repeating since it is not readily apparent from the present paper. If a load disturbance is described in terms of the temperature change which would result if the controller were inoperative, then the actual equilibrium temperature change due to a disturbance during control is $\dfrac{1}{1+s}$ times the disturbance.

One wonders why the authors introduced the auxiliary flow Q_o to effect load changes. In actual installations, load disturbances generally arise through a change in the outlet resistance. From Equation [19], it is seen that the sensitivity constant depends upon the outlet resistance. Thus the sensitivity constant normally changes with load changes. The need for characteristic flow valves, such as percentage or logarithmic valves, arises from this consideration.

In the section dealing with "Floating Control," the authors have introduced a constant c which corresponds to control band b for proportional control. In floating control, the valve travel depends upon the area of temperature deviation and time. The constant c has a real physical significance, namely, that on a time-temperature-deviation graph (which may be a temperature-controller chart) it represents the area necessary to cause complete valve travel.

Corresponding to the proportional sensitivity constant s the authors have derived a floating constant q, which may be termed "floating over-all sensitivity" and defined as the rate of potential temperature change for unit actual temperature change.

To the writer it seems desirable to describe an instrument's action in terms of the ratio of output change to input change. The narrower the control band, the greater should be the proportional effect since more valve action takes place for a given controller input change. Therefore, the proportional effect is proportional to $1/b$ and, correspondingly, the floating effect is proportional to $1/c$. On this basis the authors reset constant r, defined as the ratio b/c, should be called the ratio of the integrating to the proportional effect.

In connection with Fig. 11, it is noted that the reset constant was held unchanged and the control band decreased. It is of interest to note that, for several commercial pneumatic instruments, the mechanism is so designed that a change in the proportional adjustment affects the floating action, whereas a change in the floating adjustment does not affect the proportional action.

The authors' analysis has clarified the control effects of independent adjustments of the proportional and floating components of control. They point out that there is a critical reset rate for every process. Does this mean that there is one best floating adjustment, or one best ratio of floating-to-proportional adjustment, and how is this adjustment arrived at?

The ultimate value of this paper will depend upon its practical application. The authors have made a rational analysis of the control problem from which it is hoped valuable information will be obtained regarding guides to practical rules for industrial use. One might wish that the authors had devoted some space to the constants which arise from consideration of an actual process.

AUTHORS' CLOSURE

Messrs. Smith and Fairchild suggest that the model system shown in Fig. 1 of the paper is a relatively uncommon one. In the interest of clarity, it should be stated that no attempt was made by the authors to select an actual process for treatment. Perhaps the system selected is uncommon in that sense. However, it is both common and relevant in another and more important sense, since it involves and represents types of lag which must frequently be dealt with in the successful industrial application of control. In addition, the model system chosen has the valuable property that it submits gracefully to a relatively simple mathematical analysis, and requires a low-order linear differential equation for its complete dynamic description.

The authors have been particularly careful to avoid dealing with processes, either hydraulic or thermal, which might involve questions as to the practicability of the assumptions necessary to construct their differential equations. Under these conditions, the accuracy of the results may be relied upon to the same de-

gree as the mathematical manipulation, the latter being carried on according to accepted procedures. Until there is a thorough understanding of the behavior of nondistributed or lumped systems as such, the authors see no valid object in precarious generalizations. The possibilities inherent in the use of lumped parameters have been marvelously demonstrated over and over again in the electrical field, where highly complex networks have been represented to an uncanny degree of precision by such practical methods. In these cases, certainly, a complete knowledge of the simpler component systems was already available from previous work extending over many years. The infant science of automatic control can hardly disregard this established method of development.

The differential equation of the thermal process, of which Fig. 11 of an earlier paper[13] was taken to be an analog, was developed in that paper. Reference to that paper will show that the very reasonable assumptions which were made in the development of the thermal system will permit it to be considered as devoid of distributed capacities and resistances. Thus it is not encumbered by these complications; its thermal-lag characteristics are reasonably embodied in the structure of a second-order linear differential equation in terms of heat flow, temperature, and time.

Whether the system under control is called a "plant" or a "process" is a question which loses importance when models are under discussion. A great deal of incorrect but descriptive nomenclature has grown out of common usage in the field. This same usage has given a meaning to the word "plant" somewhat broader than would seem appropriate. The simplification obtainable in the hydraulic model by using a volume unit for quantity rather than a weight unit was foregone because of the attendant loss of vividness in the correspondences between the units of the hydraulic system and those of the various other systems which might be represented (for example, refer to Table 1 of the paper).

With regard to the "philosophy of control," the causal relationships between the manipulated and controlled variables, as pointed out in the paper, are held to be more fundamental than the simple polarity or phase criterion mentioned by Messrs. Smith and Fairchild.

Unless the term "metering lags" is interpreted as including discontinuities such as dead zones, we cannot agree that the addition of metering lag leads necessarily to a continual cycling under two-position control.

"Two-position-with-rate" control was given in the paper as a tentative name. We agree that this method of control is commonly known as "constant-speed-floating," but it was desired to emphasize its similarity to two-position control rather than to floating control. The last paragraph in the section "Two-Position-With-Rate Control" refers to a condition in which there is a predominance of capacity in that part of the process in which the controlled variable is measured. Under such a condition, the rate of corrective action could be materially increased if the variations involved in such action might be disregarded. In practice, however, dead zones are frequently employed to reduce valve activity. As pointed out by Messrs. Smith and Fairchild, this reduction in the valve variation is only accomplished by using extremely slow rates. On the other hand, if the rate of valve movement is increased indefinitely, this type approaches two-position control and the valve occupies its limiting positions practically all of the time. This was not borne out in Fig. 7 of the paper since, in none of the transients shown did the manipulated variable attain a limiting value, although it was shown in that illustration how extremely slow the rates of corrective action must be made to attain small flow variations.

It is certainly worth while to stress the physical meaning of

[13] "Quantitative Analysis of Process Lags," by C. E. Mason, Trans. A.S.M.E., vol. 60, 1938, p. 331.

the dimensionless sensitivity constant s, defined in Equation [19] of the paper. Even more generally (as pointed out in Mr. Wood's discussion), s may be identified as the ratio of an immediate change in "potential" level enacted by the controls to the original change in level which initiated such action. Unfortunately, there must be some misunderstanding about Equation [23] and/or the definitions of T_p, T_{cp}, and $(T_a)_{norm}$. In the remarks made in the paper, concerning the effects of an increase in the constant s, the term "period of cycling" was used to denote the cyclic period itself (or 1/frequency) which is invariably decreased by an increase in s. The orientation of the coordinate scales used for plotting control transients is naturally a matter of choice, but the straightforward method employed by the authors bears out graphically the 180-deg phase difference mentioned earlier in the discussion.

The constants q and c were given subordinate roles in the exposition; they were introduced in a natural manner in the development and were used to lend a logical symmetry to the equations. A joint consideration of the symbols used in Equations [19], [31], and [36] will show this symmetry.

In connection with Equations [34] to [36], the term $1/c$ cannot in any sense be considered dimensionless. The reset constant r, however, has the formal dimensions of inverse time. The fact that the value of r may be so chosen that an ultimate point stability is obtained without appreciable loss of oscillatory stability is of prime importance. Rather than being overgeneral, the statement in the paper concerning aperiodic behavior was perhaps not made general enough; further elaboration would have deviated from the main purposes of the paper.

The adjustments, demonstrated under the heading "Proportional-Plus-Floating Control," are summarized in Table 3 of the paper. Changes in the numerical value of b (or of s) have no effect on the numerical value of r, and vice versa. Their joint influence on control performance is implicit in the mathematics. The authors hope that the impression cannot be gained from the paper that the subject has had complete mathematical treatment. Such treatment has scarcely begun. Mathematics itself is only a means to an end. The progress which may be made through cooperation between user and manufacturer has been indicated by a wealth of practical evidence. Although practical experience will be invaluable in continuing the mathematical development toward some stage of practicability, it will be essential to heed the advice of Francis Bacon, given some 300 years ago: "... At the entrance of every inquiry our first duty is to eradicate any idol by which the judgment may be warped."

It is perfectly true that the original two-capacity system could have been employed. On the other hand, a more complex system might be considered necessary if a prototype, having considerable distributed resistance and capacity, were to be represented. The model chosen provided plausibly complete representation of a typical prototype involving some difficulty and, at the same time, was simple enough not to require an unwieldy analytical exposition.

The reference by Messrs. Smith and Fairchild to the work of Ivanoff is interesting and appropriate. That author's approach has been from the frequency standpoint. The two methods of treatment are equally valid and are certainly not incompatible, each having its distinct advantages. They may also serve as a check on one another in practice. The value of any method of analysis must depend upon its ability to yield laws useful in practical applications. In any case, the authors contend that there is a very real advantage in the concreteness of the representation obtainable with the liquid-level method of analogy; the more complex the system to be represented, the greater the desirability of this concreteness.

To obviate the possibility of misunderstanding, it should be

reiterated that the application of a series of controller types to the model process was made in order to demonstrate an analytic procedure found by the authors to be effective, rather than to document a search for a satisfactory controller for the particular model, as implied in the Smith-Fairchild discussion. In practice, any criterion of excellence for control is affected by economic considerations relating to the environing equipment of the controlled system concerned.

As to flow control, the general conclusions of the paper encounter no exceptions in that field. For the most part, however, automatic-flow-control problems are considered to be on a lower plane of difficulty than that dealt with in the paper, an important exception arising in those liquid-flow cases where the mass of moving fluid imparts a momentum seriously restricting the action of the controls.

Reference to the patent literature by Messrs. Smith and Fairchild is acknowledged without comment. The analysis given in the paper was directed to the mathematics of automatic control, and was not intended to refer to the particular instrument designs of any of the various manufacturers. Thus, for example, the proportional-plus-floating controller was merely postulated as a mathematical type in the development.

Mr. Wood's elaboration of the fundamental meaning of the sensitivity constant was acknowledged in a previous statement. This meaning was not directly brought out in the paper since the corresponding portion of the analysis was couched principally in terms of the empirically measurable "throttling band." In an earlier paper[13] by one of the authors, the term "potential temperature" was acknowledged to have been originated by Ivanoff.

The expression for equilibrium "loss of control point" is directly evident from Equation [20] of the paper, as

$$\Delta(T_p) = \frac{1}{1+s} \Delta(R_a Q_o)$$

where $\Delta(R_a Q_o)$ represents the load alteration whether originating in the load resistance R_a or in the load supply flow Q_o.

The mathematical method shown by the authors applies equally well to all types of disturbances. Although, for some systems, load changes are more precisely represented by changes in R_a than in Q_o, it might be questioned whether such systems are in the majority. The practical benefits arising from the use of valves having logarithmic characteristics extends far beyond the mere compensation for variation of outlet resistance with load.

Mr. Wood points out the practical significance of the constant c, thereby providing an answer to one of the questions in the Smith-Fairchild discussion.

An answer to Mr. Wood's question on the best floating adjustment may perhaps be provided by the result (indicated by the mathematical analysis of the paper) that the optimum value of the reset constant r, once it is found, is very nearly independent of the value of the throttling band which it might be advisable to employ. A controller equation faithfully describing any particular commercial instrument must be written directly from the detailed mechanics of that instrument and, thus, such an equation may not necessarily take the exact form of any of the mathematically developed types shown in the paper.

The Influence of Crystal Size on the Wear Properties of a High-Lead Bearing Metal

By JOHN R. CONNELLY,[1] BETHLEHEM, PA.

This paper,[2] describes work done to correlate metallurgical aspects of a bearing metal with mechanical testing under conditions existing in a bearing. Previous work by Karelitz and Kenyon (1),[3] and by the author (2) and (3) have concerned effects of different composition of bearing metal and different loading and size of specimen. The present investigation was made to determine the effect on removal of the bearing-metal surface and the pressure at which the oil film breaks down as a result of variations in the crystalline structure of the bearing metal. In this work both the chemical composition of the bearing metal and the size of the specimen were constant. The crystalline structure was varied and the resulting change in properties determined.

THE PROBLEM

THE analysis of bearing conditions is usually associated with the hydrodynamic theory of lubrication. This theory has served very well to explain phenomena under conditions where it is applicable, namely, where the trueness of the surfaces, the speed, the load and the fluid are such that the forces which the fluid can develop are sufficient to float the journal and prevent wear. In a number of cases, the hydrodynamic theory is not applicable, at least, as it is understood at present. The fact that most bearings show some wear after service indicates that a fluid film may not be present at all times. Thus, many bearings, which normally are running under conditions quite similar to those assumed in the hydrodynamic theory, may occasionally depart from those conditions. Such departure may be due to loss in viscosity of the lubricant either by reason of high temperature or deterioration of the lubricant.

In addition, very slow speeds or very high loads may cause such departure. Under such extreme conditions, in the presence of a lubricant, the phenomena of dry rubbing surfaces may not appear. Such extreme conditions without dry friction are often referred to as "boundary lubrication" or similar descriptive terms. Some very slow disappearance of the mating surfaces accompanies boundary lubrication. This is called wear. In brief it may be said that the conditions of boundary lubrication are slow removal of one or both mating surfaces in the presence of a lubricant but absence of a lubricant film. It was the purpose of this investigation to determine the effect on removal of the bearing-metal surface or wear and the pressure at which the oil film breaks down as a result of variation in the crystalline structure of the bearing metal.

METHOD OF INVESTIGATION

A specimen of bearing material has one side machined to a plane surface. A constant holding force, hereafter referred to as the load, presses the machined surface tangent to a rotating steel cylinder, submerged in a bath of lubricant. For these tests the force was obtained by gravity. The specimen, which can move only in a direction normal to the machined surface, wears a cylindrical groove and the contact changes from line to a progressively larger area. This wearing away continues until equilibrium conditions are established between the forces causing wear and the forces resisting wear. The load divided by the projected area at any instant is referred to as the pressure and the pressure existing at equilibrium is called the final or ultimate bearing pressure or equilibrium pressure. Rate of wear is designated as volume of bearing material removed per length of travel of a point on the surface of the cylinder.
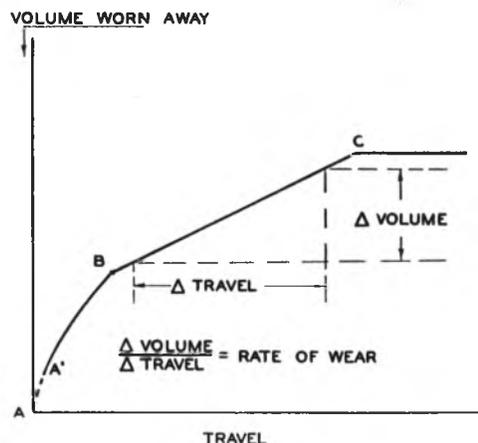
FIG. 1 DETERMINATION OF RATE OF WEAR

This wearing away of the bearing material seems to occur in more than one phase, as illustrated in Fig. 1. From A to A' results are in doubt because of the difficulty of measuring the rapid movement. Within a few minutes after the test starts, the rate of wear becomes more stable and proceeds along the path $A'B$. To distinguish $A'B$ from the following phase BC, $A'B$ is referred to as unlubricated wear, although the phrase may not be completely descriptive. Data previously analyzed show rate of wear in phase $A'B$ proportional to the corresponding contact pressure. The phase BC of the rate of wear has a linear relation with travel for all data that have come to the author's attention. This constant rate of wear, for pressures just above the ultimate bearing pressure, fixes the rate of wear in phase BC as independent of the bearing pressure. The phase BC is referred to as lubricated wear.

PREPARING TEST MATERIALS

BEARING-METAL SPECIMENS

The bearing metal selected for this investigation is known as

309

Cosmos metal and has an approximate composition of lead 76 per cent, antimony 15 per cent, and tin 9 per cent. This composition is in common use in industry. It is often used as a substitute for true babbitt metal because the high lead composition is less expensive. A further reason, for selecting this particular composition for this investigation, is that considerable variation in crystal size can be obtained by varying rates of cooling of the metal just prior to and including solidification.
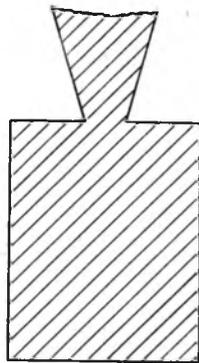
FIG. 2 TEST BAR AND RISER
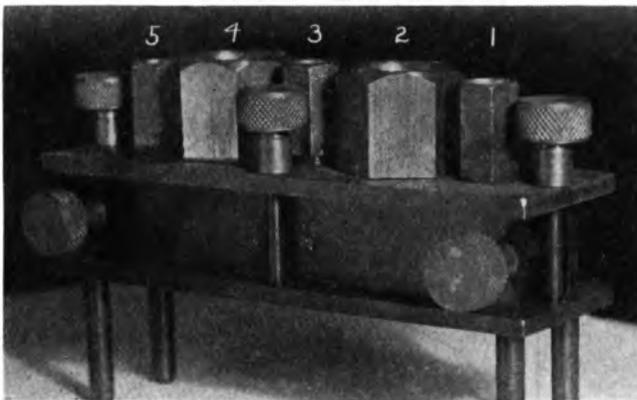
FIG. 3 MOLD FOR TEST BARS AND RISERS

A permanent mold of pieces of steel clamped together was constructed which cast a bar $1\frac{1}{4}$ in. by 1 in. by 6 in. Fig. 2 illustrates the cross section through a riser. When first made the mold had three risers which are Nos. 1, 3, and 5 in Fig. 3. Specimen bars contained cavities so risers Nos. 2 and 4 were added. Risers Nos. 1, 3, and 5 had a total volume of about 8 per cent of the bar. Risers Nos. 1, 2, 3, 4, and 5 have a column of about 31 per cent of the bar. With these additional risers no cavities appeared in the specimen bars. The risers were constructed as
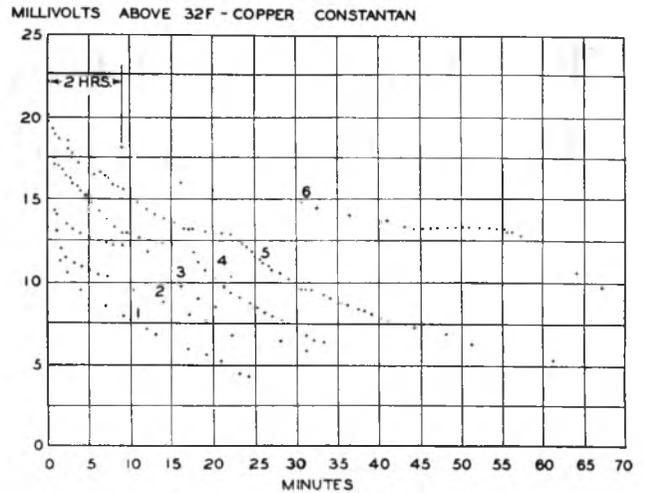
FIG. 5 COOLING OF TEST BARS
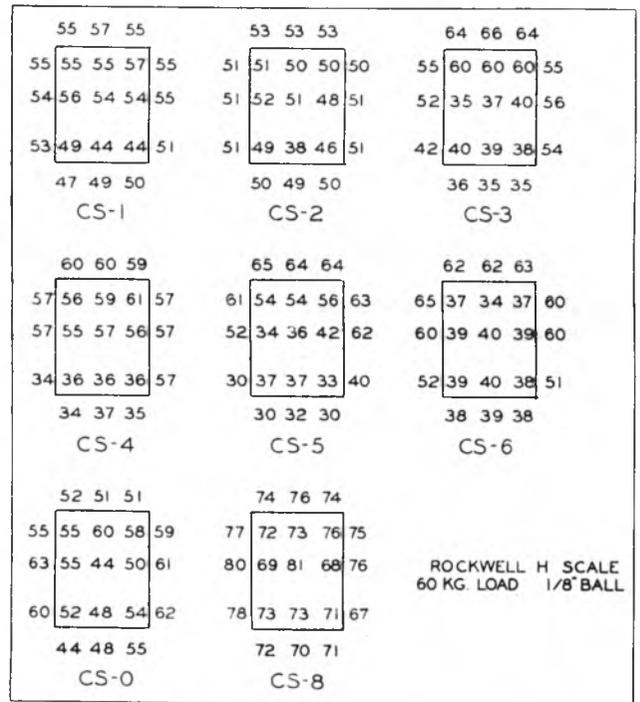
FIG. 6 HARDNESS READINGS; SECTION AND SURFACE OF TEST BARS
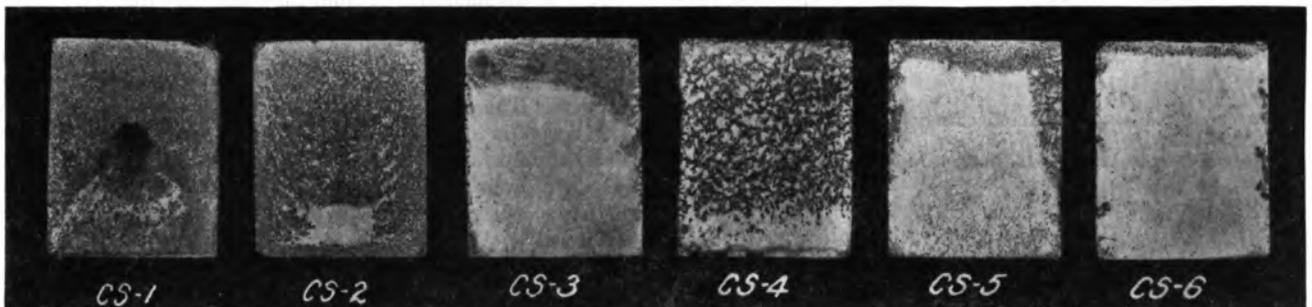
FIG. 4 MACROSTRUCTURE OF TEST BARS FOR INVESTIGATION WITH THEIR DESIGNATIONS

TABLE 1 CRYSTAL DISTRIBUTION IN SPECIMENS

Specimen	Area devoid of crystals
CS-1	4 per cent (estimated)
CS-2	8.4 per cent (by planimeter)
CS-3	71.3 per cent (by planimeter)
CS-4	Unobtainable (by planimeter)
CS-5	78.5 per cent (by planimeter)
CS-6	93.3 per cent (by planimeter)

inverted cones to assist in removing them without damage to the specimen bar. In Fig. 3, riser No. 3 is shown partially unscrewed; the unscrewing breaks the riser at the narrow section.

Six bars, hereafter designated as CS-1, 2, 3, 4, 5, and 6, were poured, Figs. 4 and 5. From these six bars were obtained (a) a broken section, (b) a polished and etched section to show microstructure, (c) a polished and overetched section to show macrostructure, (d) a section for hardness survey Fig. 6, and (e) a wear-test specimen. Of the above items (b), (c), and (e) were expected to yield the principal data. Items (a) and (d) are frequently given in metallurgical studies and for completeness were included. Also the value of that property indicated by hardness tests in this field is controversial and further information may be
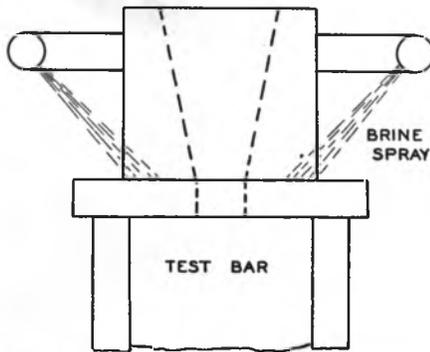
FIG. 7 METHOD OF CHILLING SPECIMEN
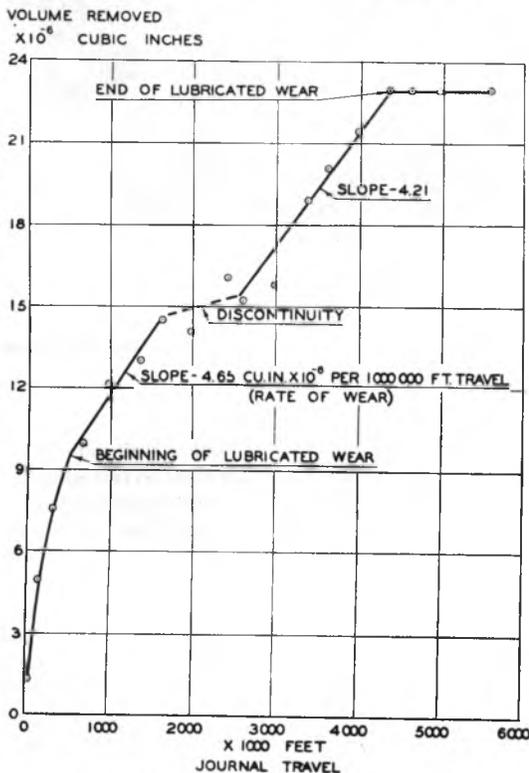
FIG. 8 DETERMINATION OF RATE OF WEAR

of value. The wear specimens were machined to shape by a new milling cutter used exclusively for this work. The $2\frac{1}{2}$-in. cutter had a face of 4 in. and 16 teeth. The test surface for each specimen was finished with the cutter rotating at 500 rpm and a linear feed of $\frac{5}{8}$ in. per min. Each wear specimen was machined to a length of approximately 0.4 in. Previous investigations indicate this length to be sufficiently beyond the range where end leakage for this type of test is a factor. Tests using the varying-wear method as described were run on these specimens.

Examination of the macrostructure of specimens CS-1, 2, 3, 4, 5, and 6, Fig. 4, shows that except for CS-4 there is a core practically devoid of crystals. The extent of crystalline structure is indicated in Table 1. This core increases in size as the rate of cooling decreases, as indicated in Table 1. The tin-antimony crystals solidify before the lead so that, when the solidification of lead is delayed, the crystals float toward the top of the bar. According to metallurgical explanation this excess of tin and antimony at the top of the bar will cause some of the antimony to crystallize as pure antimony. Antimony being a relatively hard substance, this will give a greater hardness at the surface, as

TABLE 2 RESULTS OF RATE-OF-WEAR INVESTIGATION

| Bar and test	Equilibrium press., lb per sq in.	Rate of wear, in.³ × 10⁻⁶	Avg dimension, edge of crystal, in.	Press. at start of straight-line portion of wear, lb per sq in.	Length of specimen, in.
CS-1-2t	397	4.20(5) 4.65(4)	0.0025	535	0.414
CS-1-3t	600	9.70(3)	0.0025	...	0.414
CS-1-4t	278	22.40(7)	0.0025	...	0.414
	AM 425	AM 11.80			
	GM 405	GM 9.07			
CS-2-1t	835	0.65(4)	0.0015	905	0.401
CS-2-2t	662	1.10(2)	0.0015	680	0.401
CS-2-3t	610	16.80(4)	0.0015	710	0.401a
CS-2-4b	639	1.92(5)	0.0015	...	0.401
CS-2-3b	681	2.04(4) 2.34(3) 1.70(5)			
	AM 704	AM 1.63			
	GM 700	GM 1.49			
CS-3-1t	540	2.98(4)	0.0040	610	0.393
CS-3-2t	440	5.20(4)	0.0040	460	0.393
CS-3-3t	510	6.80(6)	0.0040	585	0.393
CS-3-4t	386	10.00(4) 8.60(7)	0.0040	535	0.393
	AM 469	AM 6.95			
	GM 465	GM 6.48			
CS-4-1t	340	13.50(4) 12.10(4)	0.0030	495	0.413
CS-4-2t	304	9.20(3) 9.60(4)	0.0030	335	0.413
CS-4-3t	292	13.75(4) 17.80(7)	0.0030	370	0.413
	AM 312	AM 13.5			
	GM 311	GM 13.2			
CS-5-1t	433	2.320(5)	0.0075	510	0.413
CS-5-2t	568	0.425(4)	0.0075	600	0.413
CS-5-3t	540	1.45(6)	0.0075	780	0.413
CS-5-1b	676	1.95(5)	0.0075		
CS-5-3b	578	1.50(3) 1.50(4)			
	AM 559	AM 1.57			
	GM 553	GM 1.41			
CS-6-1t	453	6.90(3)	0.0100		0.414
CS-6-2t	523	3.50(5)	0.0100	650	0.414
CS-6-3t	387	3.50(6)	0.0100	470	0.414
	AM 454	AM 4.22			
	GM 451	GM 4.05			
CS-8-1t	604	4.00(6)	0.0060	730	0.397b
CS-8-2t	630	2.76(7)	0.0060		0.397b
	AM 617	AM 3.33			
	GM 617	GM 3.28			
CS-0-1t	638	1.64(5)	0.0010	725	0.402
CS-0-2t	603	1.13(6)	0.0010	880	0.402
CS-0-4t	980	2.48(2)	0.0010		0.402
CS-0-5t	840	2.76(4)	0.0010		0.402
	AM 765	AM 1.82			
	GM 750	GM 1.71			

NOTE: Numbers in parentheses in this table denote number of plotted points which slope value represents.

a Neglected on basis of photomicrograph.
b Different composition.
t indicates top.   AM signifies arithmetic mean.
b indicates bottom.   GM signifies geometric mean.
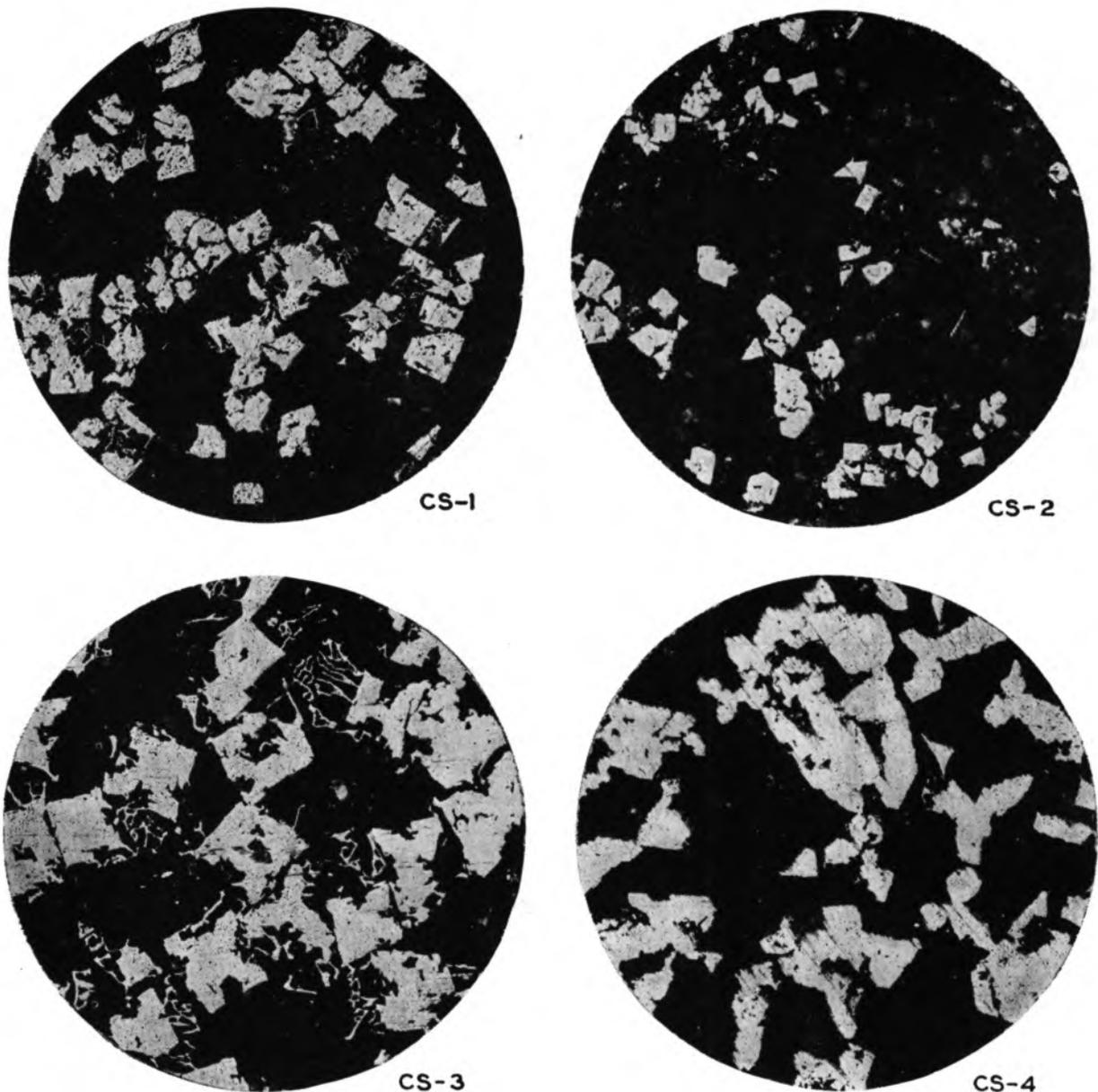
FIG. 9   TOP OF TEST BARS, MICROSTRUCTURE,   ×100

shown by the hardness survey, Fig. 6. Such hardness might affect the test results materially and obliterate the effect of crystal size. To investigate this possibility bar CS-7 was poured having reduced percentages of tin and antimony and was allowed to cool overnight in the furnace. The intention was that tin and antimony in floating to the top would give about a normal concentration and composition in that region. Bar CS-7 was so soft that it was damaged in attempting to obtain a broken section. For this reason it was discarded. A new bar CS-8 was prepared in the same manner as CS-7 and packed in dry ice for several hours. A broken section was obtained without difficulty.

It is questionable whether such a broken section is indicative of structure. It was decided to cast additional bars and attempt to secure crystals smaller than previously obtained. CS-0 was poured of the same composition as CS-1, 2, 3, 4, 5, and 6 and chilled by spraying cold brine at 15 F onto the top of the mold, as shown in Fig. 7. The bars CS-8 and 0 were given the same series of tests as were CS-1, 2, 3, 4, 5, and 6. Etching of specimens was by ferric-chloride solution or a 3 per cent solution of nitric acid in ethyl alcohol, Fig. 9.

TEST JOURNAL

The test journal was a piece of $1/2$-in-diam drill rod. No treatment was given the rod but it had been in service some months before starting these tests using the same bearing-metal-and-oil combination. This use had polished the drill rod but the change, if any, in diameter was less than 0.0001 in.

LUBRICANT

The lubricant used was D.T.E. light turbine oil secured from the Socony-Vacuum Oil Company. As received it has a Saybolt viscosity of about 169 sec at 100 F. After some use the value rises slightly. This oil had been in use for several years and had a Saybolt viscosity of 173 sec. During the tests the oil was kept at an average temperature of 77 F and fluctuated between 75 F and 80 F.
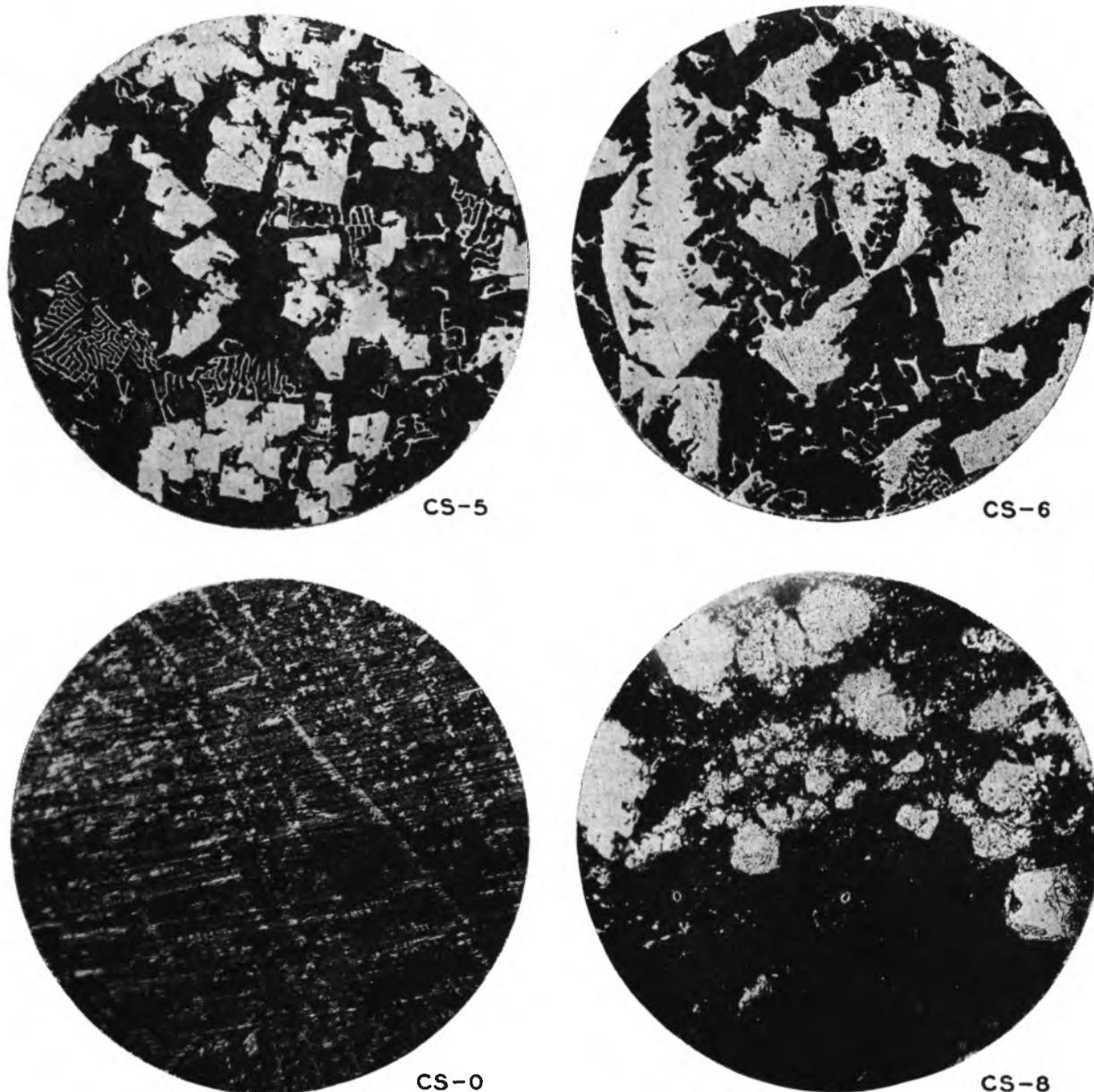
FIG. 9 (*continued*) TOP OF TEST BARS, MICROSTRUCTURE,   ×100

SUMMARY OF TEST DATA

As previously described, the rate of wear is obtained by measuring the slope of the volume curve immediately preceding the attainment of equilibrium conditions. The unit of slope is arbitrarily taken as cubic inches of volume removed per million feet of travel of a point on the surface of the journal. Fig. 8 illustrates the method of procedure. Since this bearing metal is heterogeneous, regions of variable properties occur. Discontinuities in the volume removal data evidence that the rotating journal passed through such regions. These discontinuities do not occur in every case and seem to have no regularity when present. The test results are given in Table 2. In each case, the number in parentheses, following tabular values of rate of wear, denotes the number of plotted points which the slope value represents. The larger the number of points the more reliable the rate of wear.

The dimensions of the crystals given in the table were obtained with the aid of professional metallurgists. These values repre-sent the inspection of more areas than would be feasible to present in a report.

A further item included in Table 2, showing results, is the pressure at which the change-over from unlubricated to lubricated wear takes place.

DISCUSSION OF DATA AND RESULTS

GRAPHICAL REPRESENTATION

Figs. 10, 11, 12, and 14 show results taken directly from the tabular values. Figs. 10 and 11 show equilibrium pressure and rate of wear versus the linear dimension of the tin-antimony cubical crystals. Fig. 12 presents rate of wear and equilibrium pressure against each other. Fig. 14 shows the relation between the pressures obtaining at the beginning and end of lubricated wear.

Fig. 13 shows the mass centers of results in Fig. 12. The x and y coordinates of these mass centers, referred to as $\bar{x}$ and $\bar{y}$ are

TABLE 3   COMPARISON OF RESULTS

Test bar	Crystal size		Equilibrium lb per sq in.			Rate of wear, in.³ × 10⁻⁶ per 1,000,000 ft		
	Edge, in.	Area, in.² × 10⁻³	Mass moment	Arith. mean	Geom. mean	Mass moment	Arith. mean	Geom. mean
CS-0	0.0010	0.00100	667	765	750	1.66	1.82	1.71
CS-1	0.0025	0.00625	367	425	405	11.70	11.80	9.07
CS-2	0.0015	0.00225	682	704	700	1.78	1.63	1.49
CS-3	0.0040	0.01600	428	469	465	7.71	6.95	6.48
CS-4	0.0030	0.00900	310	312	311	13.70	13.50	13.20
CS-5	0.0075	0.05625	559	559	553	1.58	1.57	1.41
CS-6	0.0100	0.10000	446	454	451	3.92	4.22	4.05
CS-8	0.0060	0.03600	618	617	617	3.31	3.33	3.28

determined by calculations analogous to those used in determining the location of the centroid of a group of rivets in a riveted joint. In place of the area of a given rivet in the analogous method, the number of slope points minus 2 is used as the multiplier for the coordinates of points shown in Fig. 12. By the number of slope points is meant the number of plotted points which the slope value represents, as described previously. Since the slope is a straight line, it is determined by two points. Subtracting 2 from the number of points which the slope represents gives the number of agreeing values in excess of those required to determine the slope line. The quantity of excess agreeing values, obtained in this way, is a conservative measure of the reliability (4) of the slope line.

Table 3 is a summation of the results showing relation between crystal size and averages of rate of wear and equilibrium pressure. The arithmetic mean, the geometric mean, and the mass centers from Fig. 12 are included. Arithmetic mean refers to the method of dividing the sum of data by the number of values. Geometric mean or logarithmic mean is obtained by taking the antilogarithm of the arithmetic mean of the logarithms of the data values. This method is employed in computations where averages are desired for data in which the nature is such that there is a lower limit to the values but no upper limit. The geometric mean is considered to give a more representative average of such data.

Figs. 15 and 16 show the averages of rate of wear against the
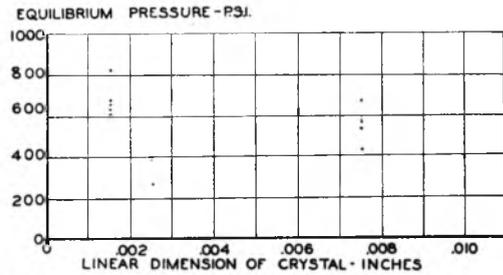
FIG. 10   EQUILIBRIUM PRESSURE VERSUS CRYSTAL SIZE; SCATTER DIAGRAM
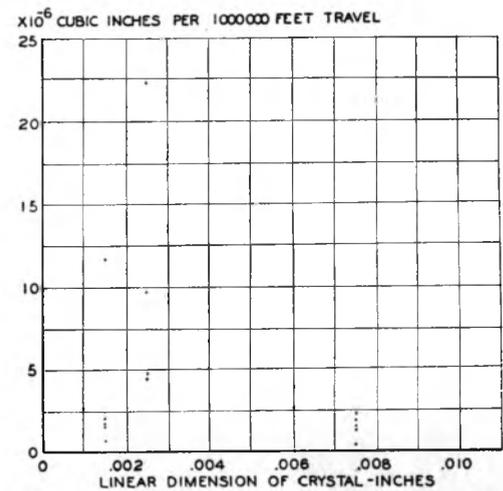
FIG. 11   RATE OF WEAR VERSUS CRYSTAL SIZE; SCATTER DIAGRAM
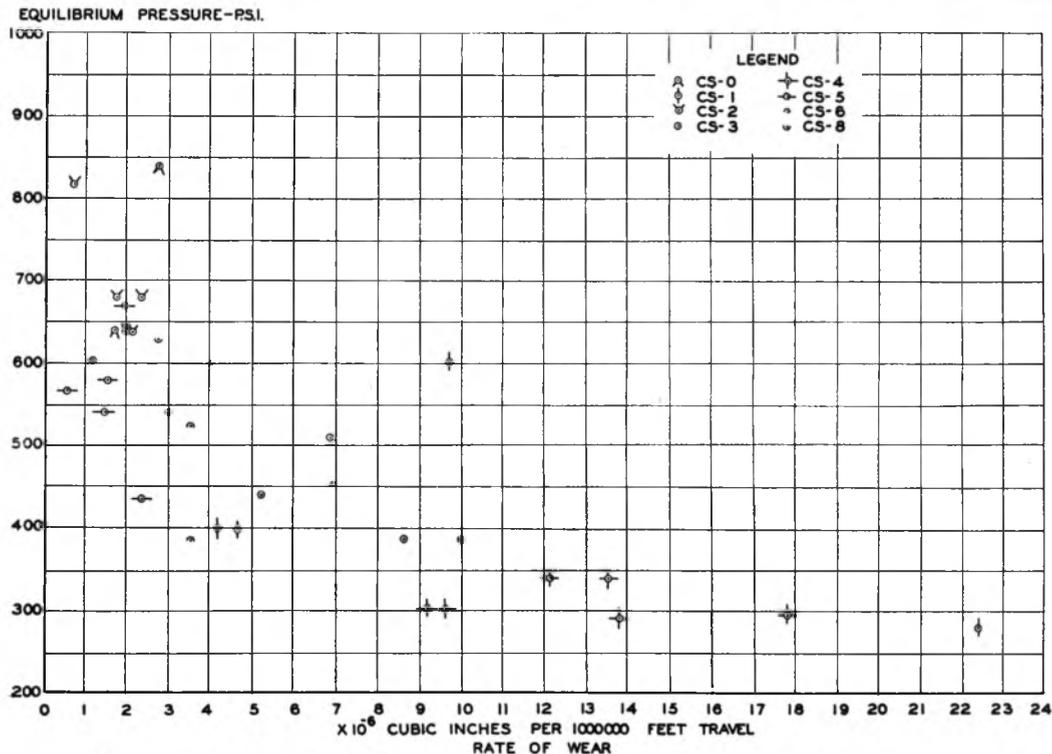
FIG. 12   EQUILIBRIUM PRESSURE VERSUS RATE OF WEAR; SCATTER DIAGRAM
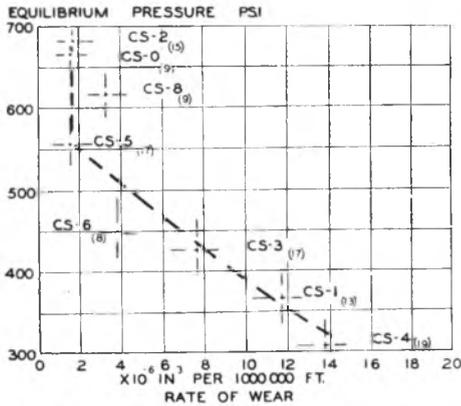
FIG. 13 EQUILIBRIUM PRESSURE VERSUS RATE OF WEAR; MASS CENTERS
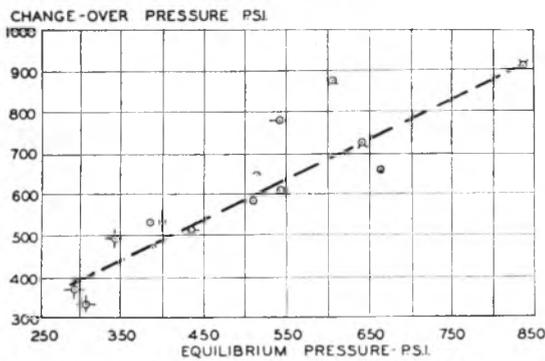
FIG. 14 EQUILIBRIUM PRESSURE VERSUS CHANGE-OVER PRESSURE
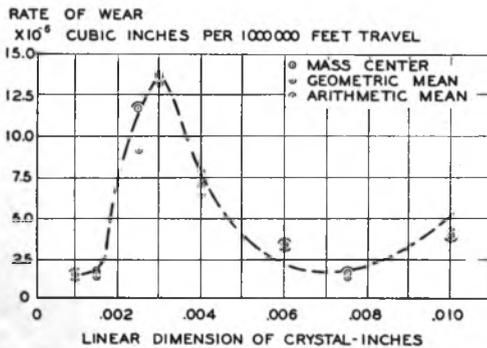
FIG. 15 AVERAGES OF RATES OF WEAR VERSUS CRYSTAL SIZE; EDGE
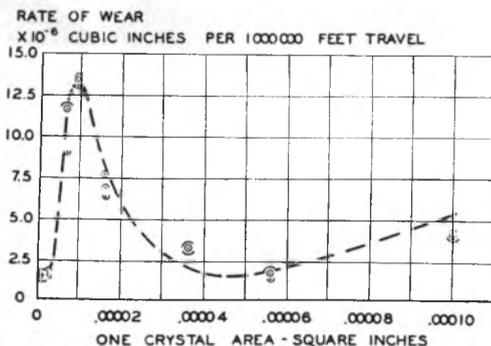
FIG. 16 AVERAGES OF RATES OF WEAR VERSUS CRYSTAL SIZE; AREA

linear dimension of the crystals and against the exposed area of one crystal that is the average crystal area. In passing a curve to represent the data, the mass centers have been favored, since the method of their determination takes account of the presumed relation between equilibrium pressure and rate of wear. The departures of samples CS-6 and CS-8 from the general trends indicated in Fig. 14 are also considered in passing the curve.

Figs. 17 and 18 represent the averages of equilibrium pressure against linear dimension of the crystals and against the average
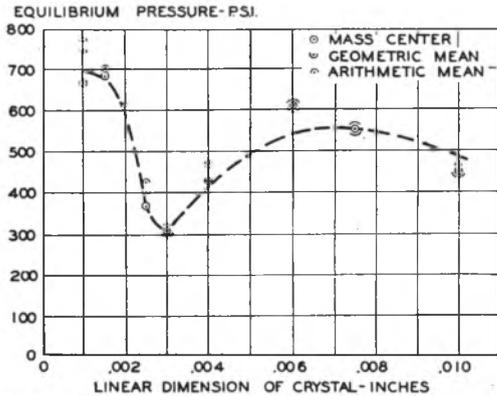
FIG. 17 AVERAGES OF EQUILIBRIUM PRESSURES VERSUS CRYSTAL SIZE; EDGE
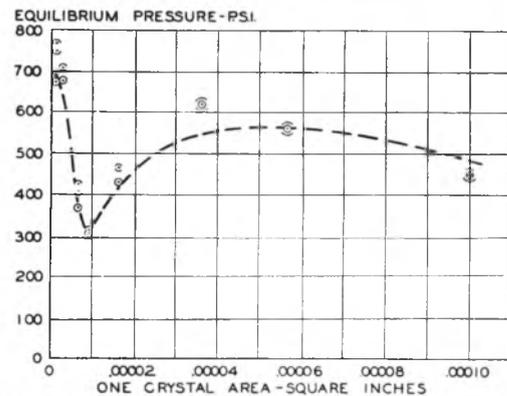
FIG. 18 AVERAGES OF EQUILIBRIUM PRESSURES VERSUS CRYSTAL SIZE; AREA

exposed crystal area. The passing of a curve to represent the data was done in the same manner as that described for Figs. 15 and 16.

COMMENTS

Whatever property of the bearing metal makes for a low rate of wear makes for a high equilibrium pressure and vice versa. On the basis of the mass centers, Fig. 13, it would seem that, below a certain crystal size, rate of wear changes little and the equilibrium pressure increases rapidly.

Fig. 12 shows that each pair of values of rate of wear and its corresponding equilibrium pressure are relatively consistent. The results are believed to give an accurate picture of the material tested, the spread of values for a given test bar being due to the heterogeneous bearing metal.

It is evident that, regardless of the averaging method employed, the general shape of the curves in Figs. 15, 16, 17, and 18 will vary little.

COMPARISON WITH OTHER INVESTIGATIONS

In a paper (1) published in 1937, Karelitz and Kenyon reported

some work by this method of testing on several compositions of bearing metal. In subsequent discussion (1) of this paper some values of rate of wear were worked out which are somewhat comparable to the data presented in the present paper. Karelitz and Kenyon report that according to their photomicrographs the specimens were overchilled. This evidence would place their data in the range of CS-0, although the various hardness values do not agree. For the lead-base specimen of Karelitz and Kenyon, the rate of wear in the same units as used in this paper is 0.92×10^{-6} cu in. per 1,000,000 ft of travel. The equilibrium pressure is approximately 700 lb per sq in. These check very well with the data here reported.

CONCLUSIONS

The results are evidence that, for the bearing metal tested, (a) crystalline structure is a factor and crystal size an important variable in thin-film lubrication, (b) rate of wear and equilibrium pressure are related, and (c) pressures at beginning and end of lubricated wear are related. The method of investigation used is a way of determining quantitative values of rate of wear, equilibrium pressure, and the pressure at the borderline condition between unlubricated wear and lubricated wear. Other things being equal a very small crystal size gives optimum values of rate of wear and equilibrium pressure.

In corroboration of the results obtained, it may be pointed out that the motorcar industry at present produces bearing inserts by a technique which gives very small crystals having a linear dimension of the order of 0.001 in. Since this is done deliberately we may infer that gross testing of over-all bearings has indicated the lower rates of wear obtained with small crystals.

Bearings that are poured in place, such as large pillow-block bearings, have crystal sizes in the range where the rate of wear and equilibrium pressure are decidedly not optimum. In the future, such large bearings could be better designed to make use of some type of insert that could be cooled rapidly upon pouring.

Of possibly greater importance than the specific results on the effect of crystal size are certain implications in connection with fundamental analysis of a journal-lubricant-bearing combination. The present concepts of thin-film lubrication take no account either of the material or crystalline structure of the mating surfaces.

In the light of this investigation, it would be appropriate to revise the phrase "hydrodynamic theory of lubrication," to be stated as the "theory of hydrodynamic lubrication." This would indicate correctly that the theory refers to a certain part of the field of lubrication and not to the entire field.

A theoretical explanation of the mechanism of wear should wait on further research in the field of thin-film lubrication. The explanation should be developed independent of the data in the range of thick-film lubrication.

ACKNOWLEDGMENTS

The author wishes to express thanks to A. D. Shankland, engineer of tests, and M. W. Dalrymple, metallurgical supervisor, of the Bethlehem Steel Company for advice and help and for photomicrography, photomacrographs, and hardness survey of crystal specimens CS-1, 2, 3, 4, 5, and 6; Professor Allison Butts of Lehigh University for advice in preparation of a crystal specimen of reduced tin-antimony content; J. H. Vail, James Ward Packard Fellow in Mechanical Engineering 1936–1938, for a major part of the data from the varying-wear machine; R. L. Scott, James Ward Packard Fellow in Mechanical Engineering 1938–1940 for the remainder of the data on the varying-wear machine; Professor F. V. Larkin, head of the department of mechanical engineering at Lehigh University, for securing funds to carry on the investigation.

BIBLIOGRAPHY

1 "Oil-Film Thickness at Transition From Semifluid to Viscous Lubrication," by G. B. Karelitz and J. N. Kenyon, Trans. A.S.M.E., vol. 59, April, 1937, pp. 239–246. Discussion, Trans. A.S.M.E., vol. 60, January, 1938, pp. 81–85.
2 "Bearing Investigation by the Varying Wear Method," by John R. Connelly and Charles C. Hertel, Miscellaneous Papers, A.S.M.E., 1935, paper no. 15
3 "A New Method of Investigating Performance of Bearing Metals," by John R. Connelly, Trans. A.S.M.E., vol. 57, Jan., 1935, paper IS-57-1, pp. 35–39.
4 "Business Statistics," by G. R. Davies and Dale Yoder, John Wiley & Sons, Inc., New York, 1937, pp. 416–421. A discussion of degrees of freedom in the analysis of reliability and significance.

Discussion

B. F. HUNTER.[4] Recently our laboratories had occasion to examine a high-lead bearing that had failed in a steam turbine for the purpose of determining the cause of failure of the bearing metal. The bearing was made in two parts each half being composed of a steel shell into which the babbitt metal was cast. In the main portion of the bearing the lining was about $1/4$ in. thick, at the edges about $1/2$ in. thick.
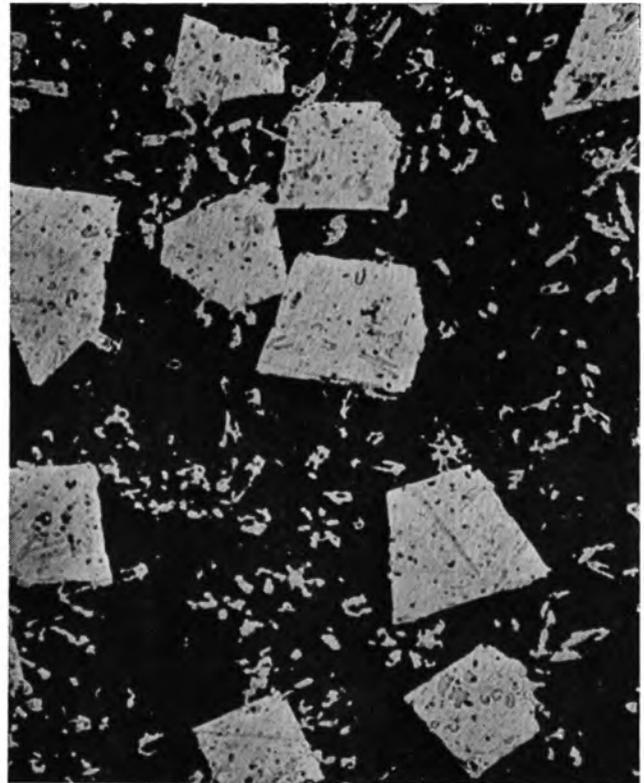
FIG. 19 REPRESENTATIVE METALLOGRAPHIC STRUCTURE OF DEFECTIVE BEARING METAL, SHOWING LARGE CUBES OF TIN-ANTIMONY, AND SHORT THICK NEEDLES OF COPPER-TIN IN COMBINATION, ×100 (ETCHED WITH FERRIC CHLORIDE)

The upper half of the bearing showed only a few slight scores but the lower half was badly scored, in addition to being cracked and pitted. Numerous small particles had fallen out and others were loose and could easily be picked out of the lining.

Chemical analysis was not made since the manufacturer of the bearing material supplied the following specifications:

Tin, per cent	84 to 86
Lead, per cent	0.035 max
Antimony, per cent	9.5 to 10.5
Copper, per cent	4.5 to 5.5

Sections from both halves of the bearing were given metallographic examinations. A representative structure is shown in Fig. 19 of this discussion. It contained extremely large cubes of Sn-Sb, and rather short thick needles of Cu-Sn in the alpha-solid solution of Sn and Cu and pseudoeutectic. At the magnification used, this latter constituent could not be resolved, but there was a considerable portion present.

The conditions attending the pouring and cooling of this bearing were not obtained and, since no analysis was made, it was assumed that the composition was correct.

That the pouring temperature was correct is revealed by the relatively short thick needles, some of which are arranged in a starlike shape. Also, the pouring temperature, except in so far as it affects the rate of cooling, is insignificant in its action toward the formation of such large cubes of Sn-Sb.

The size of these cubes, the area of each of which, as shown in Fig. 19, is about 10 times normal and their uneven distribution throughout the matrix are the only conditions present which can conceivably be the cause of the failure. The cause of these very large cubes can be attributed to the bearing alloy being too slowly cooled. The reason for the failure of the lower half was probably due to the much higher loads it had to carry; the absence of failure in the upper half was because of the comparatively light loads to which it was subjected.

The unsatisfactory service secured from this bearing appears to have been due to the presence of extremely large, poorly distributed crystals of tin-antimony compound in the bearing material, resulting from too slow a cooling rate after pouring.

This is one of many similar bearings we have examined and our findings would seem to substantiate very definitely the author's data.

F. C. Linn.[5] The author has presented an enlightening paper upon the crystalline structure and its effect upon wear and load-carrying capacity of the bearing when the operation is within the thin-film region.

Under "Conclusions" he states that "(a) crystalline structure is a factor and crystal size an important variable in thin-film lubrication." In reviewing Figs. 15 to 18, inclusive, it would appear that the linear dimension of the crystal has a very critical range, at which the rate of wear is large and the equilibrium pressure low. Outside of this range it is noted that the rate of wear is relatively low and the equilibrium pressure high. This indicates that the size of the crystal should be held within reasonable limits during the casting cycle.

It is not quite understood how the author arrived at the conclusion (c) in which he states that "pressures at beginning and end of the lubricated wear are related." A further discussion of this point would be appreciated.

The writer requests that information be supplied on the following:

1 What effect has hardness upon wear?

2 What effect has hardness on rate of wear and equilibrium pressure?

3 Referring to Fig. 6, on what surfaces were the wear tests made?

R. G. Simard.[6] The scope of this investigation on the relation

of bearing-metal wear to structural characteristics is especially commendable and, in this writer's opinion, constitutes a fruitful approach to the problem of wear phenomena. The directness of the method employed appears to have definite worth in fundamental investigations of this nature. Although there is no pretense that the work is in the category of simulated testing, the time is not considered inopportune again to remind ourselves that wear resistance is a consequence of service conditions rather than an inherent property of matter. Application to practice of the results of this or similar investigations should be attended only with extreme caution.

An investigation of the wear quality of lubricating oils in relation to representative bearing metals along lines suggested by the work of Karelitz and Kenyon at Columbia has been under way for the last 3 years in the laboratories of the research division of the Atlantic Refining Company. A battery of 6 identical machines, carefully constructed to eliminate running inaccuracies, was employed for the purpose. The standard metal specimens were prepared under rigid specification of composition and casting procedure. Results soon indicated, however, that reliance could not be attached to micrometer indication of rate of wear, since the depth of the wear groove was of such magnitude that its measurement at the end of a 5:1 lever was seriously disturbed by temperature fluctuations and slight inaccuracies of the shaft. Measurements have accordingly been confined to comparative total wear after an operating period sufficient to insure attainment of equilibrium.

The following data have been singled from our files as illustrative of that which predicated the change in procedure and these comments:

Metal	S.A.E. No. 62 bronze (hard cast)
Oil	S.A.E. No. 30 acetone-benzol dewaxed nitrobenzene refined motor oil (without film-strength additive) at 130 ± 5 F
Shaft	$\frac{3}{8}$-in. ground tap and reamer stock
Rpm	1760
Load	60 lb (nominal)

Unit	Time (days)	Measured groove width, in.	Indicated groove depth, in.	Calculated groove depth, in.
1	5	0.0139	0.0003	3.75×10^{-5}
2	7	0.0173	0.0002	5.63×10^{-5}
3	9	0.0149	0.0005	4.37×10^{-5}
4	11	0.0166	0.0004	5.63×10^{-5}
5	13	0.0154	0.0005	4.37×10^{-5}
6	13	0.0143	0.0002	3.75×10^{-5}

AUTHOR'S CLOSURE

In discussing the photomicrograph of the failed bearing, Mr. Hunter has opened up a field about which comparatively little is known. Since preparing the original paper, the author has investigated several additional methods of etching and is at present of the opinion that no one method gives a complete picture of the metallographic structure for the high-lead metal described in the paper. The electrolytic etch described by Weaver[7] has been quite instructive. Whether the same applies to the high tin metal cannot be stated, but it is entirely possible that some further critical conditions were not revealed by the etching agent employed.

Conclusion (c), "pressures at beginning and end of the lubricated wear are related," was made after study of the data shown in Fig. 14 of the paper, where the ordinate is the pressure at the beginning of lubricated wear (see Fig. 8) and the abscissa is the pressure at the end of lubricated wear. Mr. Linn raises the question of the effect of hardness on wear. Referring to Fig. 6, it is seen that CS-0, the specimen having the lowest rate of wear

[5] Turbine Engineering Department, General Electric Company, West Lynn, Mass. Mem. A.S.M.E.

[6] Research and Development Department, The Atlantic Refining Company, Philadelphia, Pa.

[7] "Type Metal Alloys," by F. E. Weaver, Journal of Institute of Metals, vol. 56, 1935, pp. 209–233.

and highest equilibrium pressure, has a low but not the lowest hardness value, while CS-8 (which can best be understood by adding 100 to all other hardness numbers), which has the lowest hardness number, is intermediate as to rate of wear and equilibrium pressure. Further examples could be given but no relation seems evident by considering Rockwell numbers. However, present methods of taking hardness may be inadequate to indicate the hardness of certain critical components of the bearing metal.

The experience of Mr. Simard concerning the measurement of groove depth does not seem directly applicable to the present paper. Karelitz and Kenyon reproduced the method reported by the author in his first paper (3) on the subject. However, at that time it was pointed out that the Ames dial had been discarded because of its many failings in favor of an electrically indicated micrometer screw. The use of a substantially different bearing metal and a resulting groove width of quite a different order of size would not seem conducive to comparable results.

The author has investigated the effect of temperature fluctuations and found that, while they are of measurable magnitude, the fluctuations resulting from the controlled oil temperatures are insufficient to vitiate the results. Further, on the basis of work done, since the original paper was prepared, the author is extremely cautious about accepting the phrase "specimens were prepared under rigid specifications of composition and casting procedure." This is no reflection on those who prepared the specimens, but is simply a warning that, present knowledge of nonferrous compounds of three components and higher being in such a state, we cannot be absolutely sure that a set of specimens is exactly as intended.

It seems that metallographic work will have to be done in parallel with work on wear of bearing materials and, as this is done, we must, as Mr. Simard suggested, be most careful of interpreting results. If others would care to work in the high-lead bearing metals, the author would gladly share his troubles.

Calculation of the Elastic Curve of a Helical Compression Spring

By H. C. KEYSOR,[1] CHICAGO, ILL.

Equations for the elastic curve of a helical compression spring are developed, without making the usual assumption of axial loading. The load is found to be eccentric, in general, being axial only under certain conditions. Elastic curves are given for both axial and eccentric loading. These curves are sinusoidal, showing that deflection along the bar is not linear, as is customarily assumed. A reversed deflection is found to occur in the closed end, back of the tip contact point, which increases the deflection of the spring as a whole, and shows that it is not correct to consider the entire closed end as inactive. An axial load is theoretically possible under certain conditions, but practically axial loading is difficult to obtain and cannot be counted on with any certainty. The effect of load eccentricity on stress is considered, and an approximate formula for estimating this effect is given. Reference is made to laboratory tests by another investigator[2] which are in good agreement with the theory here presented.

AN ANALYSIS of helical-spring deflection has been made by the author which yields three results:

1 The load is, in general, eccentric with consequent increase in stress.

2 The correct deduction for inactive end turns is shown to be approximately 1.2, based on the usual practice of taking the total turns as solid height divided by bar diameter.

3 The deflection at any point along the bar is obtained, thus permitting the elastic curve to be drawn.

The first two items have been discussed in a previous paper[3] by the author. The third item forms the subject of the present paper.

Fig. 1 shows the lower portion of a helical compression spring with ends closed and squared. Assume that the spring is compressed between parallel plane surfaces, as this is the usual condition of loading. Let P be the load. Any point on the bar, as point a', must therefore sustain the direct load P, a torque T, and a vertical bending moment M, the line of action of P being at some point Q. It is apparent that Q must be on a radial line through a, the longitudinal center point of the bar, for otherwise the forces and reactions on the top and bottom halves of the spring would not be symmetrical and the necessary condition of static equilibrium would not obtain. We do not restrict the position of point Q on the radial line through a; that

is, we consider the load to be eccentric by some variable amount e, leaving the magnitude of e to be determined by the conditions of the problem.

The load P can be replaced by two component loads, P_1 at a and P_2 diametrically opposite a. If $P_1 + P_2 = P$ and the center of pressure of P_1 and P_2 is at point Q the torque, moment, and direct load at the general point a' will not be affected. The reactions which balance P are P_3 at the tip contact point b, and $(P - P_3)$, distributed in some manner in the vicinity of point O where the bar makes contact with the loading plane. There is also a torque reaction at point O. The angular distance v between b and O is roughly 180 deg for most springs; the determination of v will be explained later.

The location of the contact point b is subject to variation due to the manner of forming the tapered end, and due also to the load, for it is evident that, with increasing load, b will move away from the tip so that the angular distance between contact points will diminish. Consideration must be given to this fact in interpreting the following analytical results in their relation to actual springs.

From the foregoing discussion, it is evident that we require the deflection equations for a helix fixed at the origin and sustaining
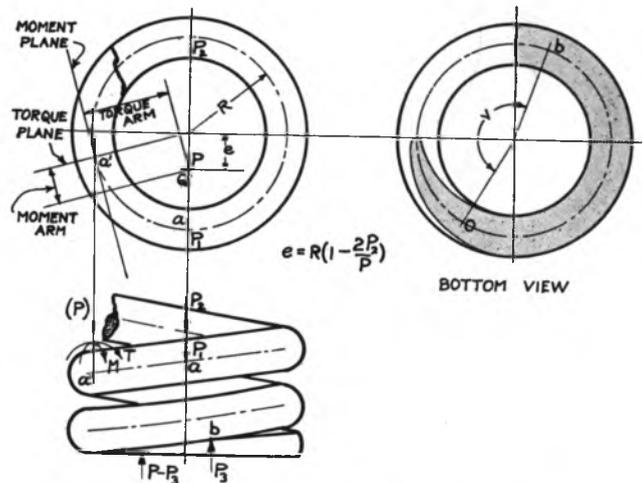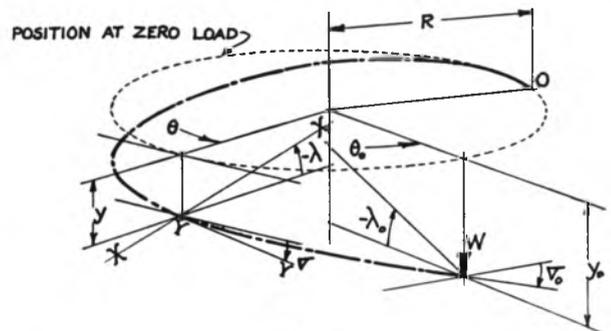

$$e = R\left(1 - \frac{2P_3}{P}\right)$$

FIG. 1

FIG. 2

---

a vertical load at any point along its length. The effects of bar curvature, direct shear, and angularity of the bar with the horizontal, are neglected, as these factors do not appreciably effect the results in the case of close-coiled springs.

Let O be the origin in Fig. 2 and W the load applied at the angle θ_0. At any point θ the deflection is y and the slopes in the circumferential plane and the radial plane are σ and λ, respectively. These three quantities, expressed as functions of θ_0 and θ, will completely define the position of the bar under load. They are obtained by the strain-energy method. It is shown in texts on elasticity that

$$y = \int \frac{Mm}{EI} dl + \int \frac{Tt}{GH} dl$$

where

M = bending moment $\Big\}$ due to applied load
T = torque
m = bending moment $\Big\}$ due to unit load applied at the point
t = torque $\quad$ whose deflection is y
EI = flexural rigidity
GH = torsional rigidity

Referring to Fig. 3, let θ be the coordinate of the point whose deflection y is required, and ψ be the coordinate of any point
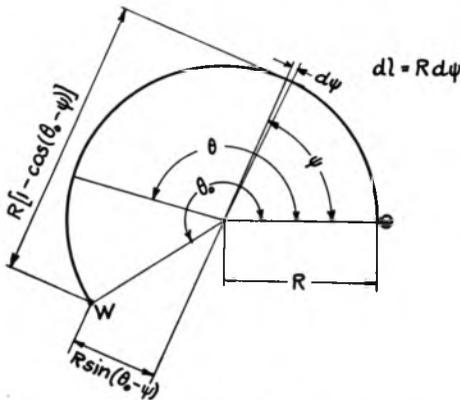
Fig. 3

between the origin O and θ. The load W is applied at θ_0, perpendicular to the plane of the figure.

At the point whose coordinate is ψ we have

$$M = WR \sin(\theta_0 - \psi)$$
$$T = WR[1 - \cos(\theta_0 - \psi)]$$
$$m = R \sin(\theta - \psi)$$
$$t = R[1 - \cos(\theta - \psi)]$$

$$y = WR^3 \left\{ \frac{1}{GH} \int_0^\theta [1 - \cos(\theta_0 - \psi)][1 - \cos(\theta - \psi)]d\psi \right.$$
$$\left. + \frac{1}{EI} \int_0^\theta \sin(\theta_0 - \psi)\sin(\theta - \psi)d\psi \right\}$$

Reducing, we obtain

$$y = WR^3 \left\{ \frac{1}{GH} \left[\theta + \frac{1}{2} \theta \cos(\theta_0 - \theta) + \sin(\theta_0 - \theta) - \sin \theta_0 \right.\right.$$
$$\left. - \sin \theta + \frac{1}{2} \cos \theta_0 \sin \theta \right] + \frac{1}{EI} \left[\frac{1}{2} \theta \cos(\theta_0 - \theta) \right.$$
$$\left.\left. - \frac{1}{2} \cos \theta_0 \sin \theta \right] \right\} \ldots \ldots [1]$$

The circumferential slope σ is found by differentiating Equation [1] with respect to $Rd\theta$ since

$$\sigma = \frac{dy}{dl} = \frac{1}{R} \frac{dy}{d\theta}$$

$$\sigma = WR^2 \left\{ \frac{1}{GH} \left[1 + \frac{1}{2} \theta \sin(\theta_0 - \theta) - \frac{1}{2} \sin \theta_0 \sin \theta - \cos \theta \right] \right.$$
$$\left. + \frac{1}{EI} \left[\frac{1}{2} \theta \sin(\theta_0 - \theta) + \frac{1}{2} \sin \theta_0 \sin \theta \right] \right\} \ldots \ldots [2]$$

The radial slope λ is shown in Fig. 2, and it is apparent that the originally horizontal radial axis X—X will slope upward toward the coil axis, while the circumferential Y—Y axis slopes downward. Therefore, if we take σ in Equation [2] as positive, λ will be negative. Let $d\phi$ be the angle of twist in the element dl at ψ, and di be the change in slope due to vertical bending, in the length dl at ψ. Projecting $d\phi$ and di upon the plane of θ, which makes the angle $\theta - \psi$ with the plane of ψ, we obtain

$$d\lambda = d\phi \cos(\theta - \psi) - di \sin(\theta - \psi)$$

It can be seen that the torque T which causes $d\phi$, tends to give a positive value of λ at θ, but the moment M would make $d\lambda$ negative. Since

$$d\phi = \frac{Tdl}{GH} \quad \text{and} \quad di = \frac{Mdl}{EI}$$

then

$$\lambda = WR^2 \left\{ \frac{1}{GH} \int_0^\theta \cos(\theta - \psi)[1 - \cos(\theta_0 - \psi)]d\psi \right.$$
$$\left. - \frac{1}{EI} \int_0^\theta \sin(\theta - \psi) \sin(\theta_0 - \psi)d\psi \right\}$$

which reduces to

$$\lambda = -WR^2 \left\{ \frac{1}{GH} \left[\frac{1}{2} \theta \cos(\theta_0 - \theta) - \sin \theta + \frac{1}{2} \cos \theta_0 \sin \theta \right] \right.$$
$$\left. + \frac{1}{EI} \left[\frac{1}{2} \theta \cos(\theta_0 - \theta) - \frac{1}{2} \cos \theta_0 \sin \theta \right] \right\} \ldots [3]$$

The deflections and slopes at the load point θ_0 are obtained by making $\theta = \theta_0$

$$y_0 = WR^3 \left\{ \frac{1}{GH} \left[\frac{3}{2} \theta_0 - 2 \sin \theta_0 + \frac{1}{4} \sin 2\theta_0 \right] \right.$$
$$\left. + \frac{1}{EI} \left[\frac{1}{2} \theta_0 - \frac{1}{4} \sin 2\theta_0 \right] \right\} \ldots \ldots [1a]$$

$$\sigma_0 = WR^2 \left\{ \frac{1}{GH} \left[\frac{3}{4} - \cos \theta_0 + \frac{1}{4} \cos 2\theta_0 \right] \right.$$
$$\left. + \frac{1}{EI} \left[\frac{1}{4} - \frac{1}{4} \cos 2\theta_0 \right] \right\} \ldots \ldots [2a]$$

$$\lambda_0 = -WR^2 \left\{ \frac{1}{GH} \left[\frac{1}{2} \theta_0 - \sin \theta_0 + \frac{1}{4} \sin 2\theta_0 \right] \right.$$
$$\left. + \frac{1}{EI} \left[\frac{1}{2} \theta_0 - \frac{1}{4} \sin 2\theta_0 \right] \right\} \ldots \ldots [3a]$$

These equations can be checked by observing that they must satisfy the condition

$$y_0 = y + \bar{y} + R\sigma \sin(\theta_0 - \theta) + R\lambda[1 - \cos(\theta_0 - \theta)] \ldots [4]$$

where $\bar{y}$ is the deflection at θ_0 referred to θ as an origin, and is obtained from Equation [1a] by replacing θ_0 with $(\theta_0 - \theta)$. Reduction of the right-hand member of Equation [4] shows that it becomes identical with Equation [1a], thus proving that Equations [1] to [3] are correct.

The deflection and slopes of a point beyond the load point are required as shown in Fig. 4.

POSITION AT ZERO LOAD
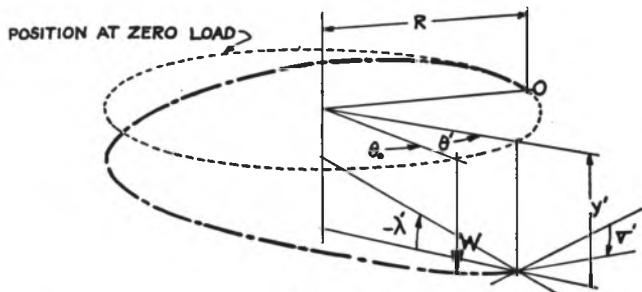
FIG. 4

Letting θ' be the coordinate of the desired point we obtain y' from Equation [4] by making $\bar{y} = 0$, $y_0 = y'$, $\theta_0 = \theta'$, and $\theta = \theta_0$:

$$y' = y_0 + R\sigma_0 \sin(\theta' - \theta_0) + R\lambda_0[1 - \cos(\theta' - \theta_0)]$$

Projecting the orthogonal slopes σ_0 and λ_0 on the plane of θ', that is, through the angle $(\theta' - \theta_0)$, we obtain

$$\sigma' = \sigma_0 \cos(\theta' - \theta_0) - \lambda_0 \sin(\theta' - \theta_0)$$

$$\lambda' = \lambda_0 \cos(\theta' - \theta_0) - \sigma_0 \sin(\theta' - \theta_0)$$

Substitute y_0, σ_0, and λ_0 from Equations [1a], [2a], and [3a]; reducing we get

$$y' = WR^3 \left\{ \frac{1}{GH}\left[\theta_0 + \frac{1}{2}\theta_0 \cos(\theta' - \theta_0) + \sin(\theta' - \theta_0) \right.\right.$$
$$\left. - \sin\theta' - \sin\theta_0 + \frac{1}{2}\cos\theta' \sin\theta_0 \right]$$
$$\left. + \frac{1}{EI}\left[\frac{1}{2}\theta_0 \cos(\theta' - \theta_0) - \frac{1}{2}\cos\theta' \sin\theta_0 \right] \right\} \dots\dots [5]$$

$$\sigma' = WR^2 \left\{ \frac{1}{GH}\left[\frac{1}{2}\theta_0 \sin(\theta' - \theta_0) + \frac{3}{4}\cos(\theta' - \theta_0) \right.\right.$$
$$\left. - \cos(2\theta_0 - \theta') + \frac{1}{4}\cos(3\theta_0 - \theta') \right] + \frac{1}{EI}\left[\frac{1}{2}\theta_0 \sin(\theta' - \theta_0) \right.$$
$$\left.\left. + \frac{1}{4}\cos(\theta' - \theta_0) - \frac{1}{4}\cos(3\theta_0 - \theta') \right] \right\} \dots\dots [6]$$

$$\lambda' = -WR^2 \left\{ \frac{1}{GH}\left[\frac{1}{2}\theta_0 \cos(\theta' - \theta_0) + \sin(\theta' - \theta_0) - \sin\theta' \right.\right.$$
$$\left. + \frac{1}{2}\cos\theta' \sin\theta_0 \right] + \frac{1}{EI}\left[\frac{1}{2}\theta_0 \cos(\theta' - \theta_0) \right.$$
$$\left.\left. - \frac{1}{2}\cos\theta' \sin\theta_0 \right] \right\} \dots\dots [7]$$

The method of applying the equations to the calculation of the elastic curve will now be given. Fig. 5 shows the lower half of the spring, a being the longitudinal center of the bar. The angle $u - v$ is known, since it is one half the angle between tip contact points. Point O, which is taken as the origin, is the same as in Fig. 1, and its position must be left to be determined, that is, angle v is an unknown. The two load components, P_1

and P_2, and the reaction P_3, are all unknown, so that, with v, we have four quantities to determine. The four necessary conditions are:

1 $P_1 + P_2 = P$ where P is the given load on the spring.

2 The combined deflection at point b is zero, since the tip prevents downward movement of the adjacent coil.

3 The combined radial slope at point a must be zero to suit the imposed condition of parallel loading planes.

4 The combined vertical bending moment at point O must be zero, for the bar is not fastened to the loading plane and is, therefore, free to move upward, which it actually does. Suppose that at some point O' between O and b the bar were forced back to its no-load position by a force F. The reaction to F would be upward at b resulting in a vertical moment at O', which point would be the origin. Therefore, a moment at the origin would necessitate a downward force at that point to hold the bar against the plane, obviously a physical impossibility. It will be shown later that the torque at the origin is not zero, and that this condition is consistent with a zero bending moment.

The combined circumferential slope at a is not required, for whatever value it may have, the condition of parallel loading planes will be satisfied due to the symmetry of the top and bottom halves of the spring.
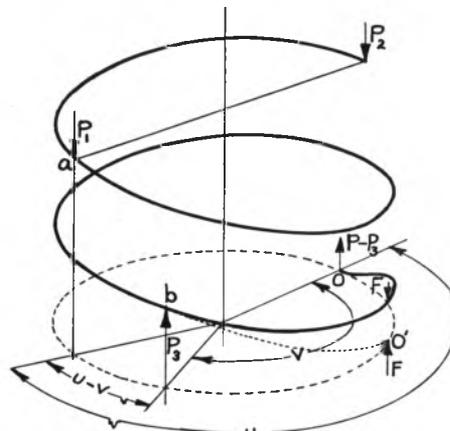
FIG. 5

To express the conditions (2) and (3), and to find the deflection at a, the component deflections are written from Equations [1] to [7]. Taking P_1 at a, we have $\theta_0 = u$ and $\theta = v$. From Equation [1a]

$$(y_a)_1 = P_1 R^3 \left\{ \frac{1}{GH}\left[\frac{3}{2}u - 2\sin u + \frac{1}{4}\sin 2u \right] \right.$$
$$\left. + \frac{1}{EI}\left[\frac{1}{2}u - \frac{1}{4}\sin 2u \right] \right\}$$

From Equation [1]

$$(y_b)_1 = P_1 R^3 \left\{ \frac{1}{GH}\left[v + \frac{1}{2}v\cos(u - v) + \sin(u - v) \right.\right.$$
$$\left. - \sin u - \sin v + \frac{1}{2}\sin v \cos u \right] + \frac{1}{EI}\left[\frac{1}{2}v\cos(u - v) \right.$$
$$\left.\left. - \frac{1}{2}\sin v \cos u \right] \right\}$$

From Equation [3a]

$$(\lambda_a)_1 = -P_1 R^2 \left\{ \frac{1}{GH}\left[\frac{1}{2}u - \sin u + \frac{1}{4}\sin 2u \right] \right.$$
$$\left. + \frac{1}{EI}\left[\frac{1}{2}u - \frac{1}{4}\sin 2u \right] \right\}$$

Taking P_2 diametrically opposite a, we have $\theta_0 = u + \pi$ and $\theta = u$ for point a. From Equation [1]

$$(y_a)_2 = P_2 R^3 \left\{ \frac{1}{GH} \left[\frac{1}{2} u - \frac{1}{4} \sin 2u \right] \right.$$
$$\left. + \frac{1}{EI} \left[-\frac{1}{2} u + \frac{1}{4} \sin 2u \right] \right\}$$

From Equation [3]

$$(\lambda_a)_2 = -P_2 R^2 \left\{ \frac{1}{GH} \left[-\frac{1}{2} u - \sin u - \frac{1}{4} \sin 2u \right] \right.$$
$$\left. + \frac{1}{EI} \left[-\frac{1}{2} u + \frac{1}{4} \sin 2u \right] \right\}$$

For point b, $\theta = v$ and we have from Equation [1]

$$(y_b)_2 = P_2 R^2 \left\{ \frac{1}{GH} \left[v - \frac{1}{2} v \cos(u-v) - \sin(u-v) + \sin u \right. \right.$$
$$\left. - \sin v - \frac{1}{2} \sin v \cos u \right]$$
$$\left. + \frac{1}{EI} \left[-\frac{1}{2} v \cos(u-v) + \frac{1}{2} \sin v \cos u \right] \right\}$$

Taking P_3 at point b and using the negative sign because P_3 acts upward, we have $\theta_0 = v$ and $\theta' = u$. From Equation [5]

$$(y_a)_3 = -P_3 R^3 \left\{ \frac{1}{GH} \left[v + \frac{1}{2} v \cos(u-v) + \sin(u-v) \right. \right.$$
$$\left. - \sin u - \sin v + \frac{1}{2} \sin v \cos u \right] + \frac{1}{EI} \left[\frac{1}{2} v \cos(u-v) \right.$$
$$\left. \left. - \frac{1}{2} \sin v \cos u \right] \right\}$$

From Equation [7]

$$(\lambda_a)_3 = P_3 R^2 \left\{ \frac{1}{GH} \left[\frac{1}{2} v \cos(u-v) + \sin(u-v) - \sin u \right. \right.$$
$$\left. + \frac{1}{2} \sin v \cos u \right] + \frac{1}{EI} \left[\frac{1}{2} v \cos(u-v) - \frac{1}{2} \sin v \cos u \right] \right\}$$

From Equation [1a]

$$(y_b)_3 = -P_3 R^3 \left\{ \frac{1}{GH} \left[\frac{3}{2} v - 2 \sin v + \frac{1}{4} \sin 2v \right] \right.$$
$$\left. + \frac{1}{EI} \left[\frac{1}{2} v - \frac{1}{4} \sin 2v \right] \right\}$$

Combining the above results we have:

$$(y_b)_1 + (y_b)_2 + (y_b)_3 = 0 \ldots \text{Condition (2)}$$
$$(\lambda_a)_1 + (\lambda_a)_2 + (\lambda_a)_3 = 0 \ldots \text{Condition (3)}$$
$$y_a = (y_a)_1 + (y_a)_2 + (y_2)_3$$

This last equation is not required for the elastic-curve calculation, but is used in calculating the deflection of the spring as a whole and, hence, the deduction for inactive end turns. The condition equations can be simplified by the usual trigonometric transformations, and for round steel bars, a further simplification is made by putting $\dfrac{1}{EI} = 0.770 \dfrac{1}{GH}$.

Simultaneous solution of the equations for conditions (1), (2), and (3) gives the results

$$P_1 = P \frac{BF + EC}{F(A + B) - 2C^2}$$
$$P_2 = P - P_1 \left. \right\} \quad \ldots \ldots \ldots \ldots \ldots \ldots [8]$$
$$P_3 = P \frac{E(A + B) + 2BC}{F(A + B) - 2C^2}$$

$$y_a = \frac{PR^3}{GH} \left[K \frac{P_1}{P} + L \frac{P_2}{P} - D \frac{P_3}{P} \right]$$

$A = 0.885u - \sin u + 0.0575 \sin 2u$

$B = 0.885u + \sin u + 0.0575 \sin 2u$

$C = 0.885v \cos(u-v) + \sin(u-v) - \sin u$
$\quad + 0.115 \sin v \cos u$

$D = C + v - \sin v$

$E = -C + v - \sin v$

$F = 1.885v - 2 \sin v + 0.0575 \sin 2v$

$K = 1.885u - 2 \sin u + 0.0575 \sin 2u$

$L = 0.115u - 0.0575 \sin 2u$

The angle v is now determined from condition (4) which is

$$\left[BF + EC - \frac{1}{2} F(A + B) + C^2 \right] \sin u - \left[\frac{1}{2} E(A + B) \right.$$
$$\left. + BC \right] \sin v = 0 \ldots \ldots \ldots [9]$$

This equation can be solved by trial. If we assign a numerical value to v, the terms containing only trigonometric functions of u can be grouped into one member, and the other member of the equation becomes merely a constant times u. To illustrate, let $v = 188$ deg, and the condition (4) equation becomes

$$\sin^2 u - 0.5494 \sin u [2.9039 \cos(u - 8°) + \sin(u - 8°) + \sin u$$
$$+ 0.0160 \cos u] + 0.004226 \sin 2u = -0.06502 u$$

The left-hand member contains only sine and cosine functions of u and its graph will therefore be sinusoidal with waves of constant amplitude. The right-hand member is merely an inclined straight line, and since it is negative, we need consider only the negative portions of the sinusoidal curve. This curve repeats at intervals of π, therefore only the first negative wave is required, and we obtain the roots by drawing a series of inclined parallel lines separated by a vertical distance of 0.06502π as

POINT NO.	ROOT
3	1.502π
4	2.080π
5	2.459π
6	3.141π
7	3.411π
8	4.221π
9	4.333π
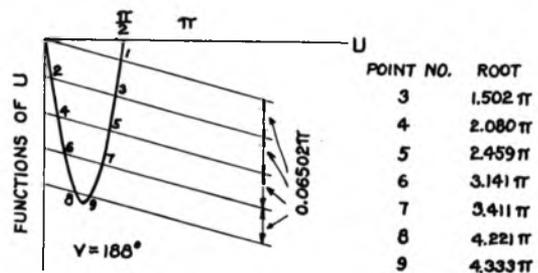
Fig. 6

shown in Fig. 6. The first two roots, points (1) and (2), have no significance since they are less than v. The other seven roots are as shown.

Values of u and v obtained in this manner are plotted in Fig. 7. It is seen that v fluctuates about 180 deg as a mean and that, except for short springs, it differs but little from this value.

We can now work out any desired numerical case from Equations [8] and the graph of Fig. 7. It is found, in general, that $P_1 \neq P_2$ or the resultant load P is eccentric. The loading is axial, $P_1 = P_2$ only when

$$u - v = \text{integer} \times \pi + 92 \text{ deg}$$

and the eccentricity is maximum when

$$u - v = \text{integer} \times \pi \text{ (approximately)}$$

The maximum eccentricity diminishes with increase of u, as shown in Fig. 7. The abscissas are shown in terms of u and also the number of turns between tip contact points.
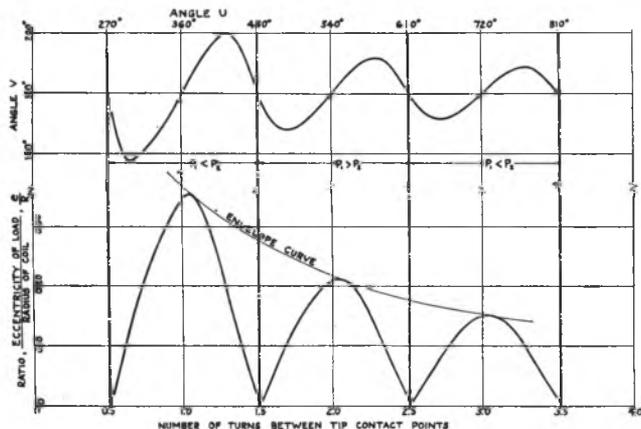
FIG. 7

The reactions between the coil end and the loading plane must now be considered, to show that the necessary conditions of static equilibrium are satisfied. The tapered tip which contacts the loading plane is shown in Fig. 8.
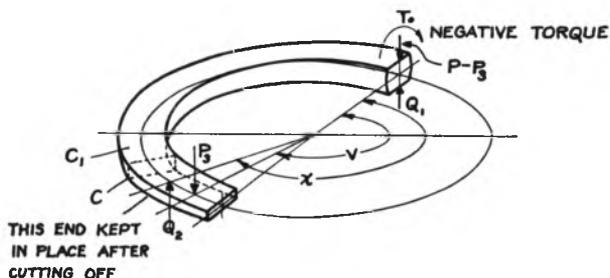
FIG. 8

A numerical calculation of P_1, P_2, and P_3 for various values of u shows that there is always a negative torque T_0 at the origin, as shown in Fig. 8, which would cause the tapered end to rise from the loading plane if it were not held down by the reaction P_3. The reactions between the tapered end and the loading plane are Q_1 and Q_2. Static equilibrium requires that the sum of torques, and the sum of bending moments at O must each be zero, and sum of reactions equal to the load on the spring:

$$P_3R(1 - \cos v) - Q_2R(1 - \cos x) - T_0 = 0$$
$$P_3R \sin v - Q_2R \sin x = 0$$
$$Q_1 + Q_2 = P$$

The terms P_1, P_2, and P_3 are calculated as previously explained, and T_0 from the equation

$$T_0 = R[- P_3(1 - \cos v) + P - (P_1 - P_2) \cos u]$$

Q_1, Q_2, and x can then be found from the foregoing three equations. The results of a sample calculation are as follows:

u	= 387 deg 7 min		
v	= 193 deg	x	= 193 deg 55 min
P_1	= 0.32407 P	Q_1	= 0.33362 P
P_2	= 0.67593 P	Q_2	= 0.66638 P
P_3	= 0.71284 P		
T_0	= —0.09423 PR		

To test the truth of the foregoing reasoning, a test was made, cutting off the tapered ends at C (in Fig. 8) and using a spring with ends ground accurately parallel (to within 0.0005 in. in $6^1/_2$ in.), and employing ground-steel plates for loading planes. Upon application of the load it was found that the end C_1 moved upward from the loading plane, thus proving that there is a negative torque at O if the ends are not cut.

The elastic curve can now be calculated for any desired length of spring. Taking first a case of approximately maximum eccentricity ($u = 1080$ deg and $v = 180$ deg) we obtain from Equation [8]

$$P_1 = 0.4520 \, P$$
$$P_2 = 0.5480 \, P$$
$$P_3 = 0.5755 \, P$$

Letting θ be the coordinate of any point, we write the component deflections from Equation [1] using

$$\theta_0 = \begin{cases} 1080 \text{ deg for } P_1 \\ 1260 \text{ deg for } P_2 \\ 180 \text{ deg for } P_3 \end{cases} \text{ and } 0 < \theta < 180 \text{ deg}$$

Adding these deflections, in terms of P, we obtain

$$y = \frac{PR^3}{GH} \times 0.4245 \ [\theta(1 + \cos \theta) - 2 \sin \theta] \ \ldots \ [10]$$

which gives the elastic curve for points between the origin O and the tip contact point b.

For values of $\theta > v$ we require another equation because P_3 now comes between θ and the origin. Using Equation [5], and replacing θ' with θ, and θ_0 with v (180 deg) we obtain the component deflection for P_3. The component deflections for P_1 and P_2 are the same as used in the preceding paragraph. Combining

$$y = \frac{PR^3}{GH} \ [\theta(1 - 0.08489 \cos \theta) - 1.8081$$
$$+ 1.6175 \sin (\theta + 81°37')] \ldots [11]$$

which applies between the tip contact b and the bar center point a.

The consistency of these results can be seen by noting that Equation [10] gives $y = 0$ for $\theta = 0$ or $\theta = 180$ deg, and Equation [11] gives $y = 0$ for $\theta = 180$ deg. Also the derivatives of Equations [10] and [11], taken with respect to θ, are equal for $\theta = 180$ deg, thus satisfying the necessary condition of continuity.

The graph of Equations [10] and [11] is the required elastic curve and is shown in Fig. 9, deflections being plotted in terms of the deflection of the entire spring. It is seen that the deflection is reversed between the origin and tip contact point, which fact makes the deflection of the whole spring more than it would be if the bar were held fixed at the tip contact. The elastic curve has sinusoidal waves, the amplitude of which diminishes as the center of the spring is approached.

Taking the case of axial loading, we have

$$P_1 = P_2 = 0.5 \, P \qquad u = \text{integer} \times \pi + 92 \text{ deg}$$
$$P_3 = 0.5305 \, P \qquad v = 180 \text{ deg}$$

The method of calculation is the same as for the previous case, and we find

$$y = \frac{PR^3}{GH} \times 0.4695 \ [\theta(1 + \cos \theta) - 2 \sin \theta] \ \ldots \ [12]$$

$$0 < \theta < v$$

$$y = \frac{PR^3}{GH} \left[\theta - 1\frac{2}{3} + 1.4763 \sin (\theta + 87°38') \right] \ldots [13]$$

$$v < \theta < u$$

Equations [10] and [12] are identical except for a small difference in the numerical coefficient, showing that the deflection back of the tip-contact point is practically unaffected by load eccen-
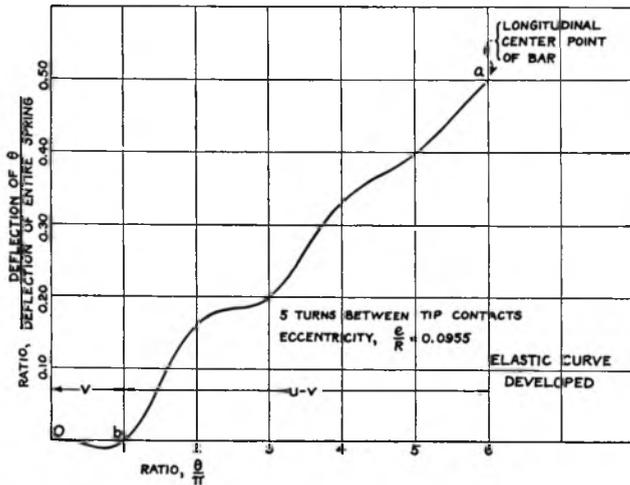
FIG. 9

tricity. The graph is shown in Fig. 10; it has sinusoidal waves of constant amplitude as can be seen from the form of Equation 13, whereas, with eccentric loading the amplitude diminishes at the bar center point. This is the distinguishing feature between axial and eccentric loading as far as the elastic curve is concerned.

An experimental determination of the elastic curve, covered in a current paper[2] by D. H. Pletta and F. J. Maher of the Virginia Polytechnic Institute, furnishes proof of the essential correctness of the analytical results.

Load eccentricity increases stress because the maximum torque is increased, the direct load being unchanged. If we let Z be the eccentricity stress factor, the relation between Z and the eccentricity e is given by

$$e = 1.123 \, R(Z - 1) \quad \text{or} \quad Z = 1 + \frac{e}{1.123 \, R}$$

Axial loading is theoretically possible, but actually it cannot be realized. This point is discussed in more detail in Professor Pletta's paper.[2] His experiments showed close agreement between the test and theoretical elastic curves in cases of maximum load eccentricity. For the theoretical number of turns corresponding to axial loading, the agreement between test and theory was not as close, indicating that axial loading was not realized in the tests in spite of all the precautions taken to obtain it. In actual springs the chances of getting axial loading would be even less, and it would therefore appear best for conservative design to calculate the eccentricity stress factor Z from the envelope of the e curves shown in Fig. 7. It can be shown that on this basis the envelope formula for Z would be

$$\bar{Z} = 1 + \frac{0.5043}{N} + \frac{0.1213}{N^2} + \frac{2.0584}{N^3}$$

N = solid ht ÷ bar diameter

In order to use the foregoing analytical results in calculation, it is necessary to relate the angle $u - v$ to the number of turns in the spring. The actual average location of the tip-contact point b, as shown in Fig. 1, is somewhat advanced from the extreme tip, say about $1/20$ turn or 18 deg. Since the bar-tip thickness is $1/4$ of the bar diameter, the number of closed turns at one end is
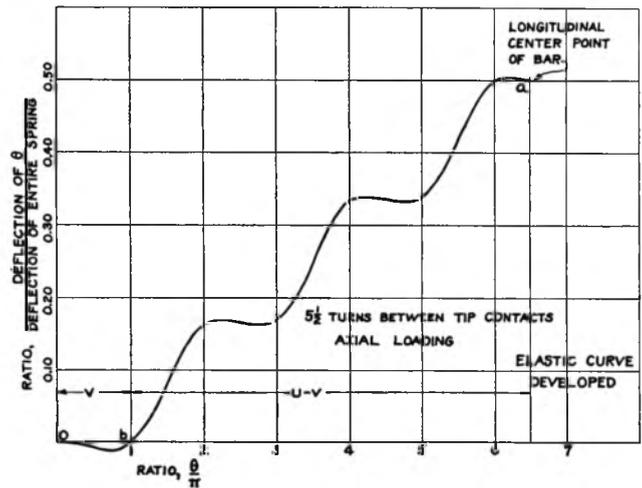
FIG. 10

$1/2 + 1/4 + 1/20 = 0.8$, or 1.6 for the whole spring. Therefore

$$2(u - v) = 2\pi(N - 1.6)$$

$$\frac{u - v}{\pi} = N - 1.6 = \text{number of turns between tip-contact points.}$$

Using this equation in connection with any desired value of N, we can find the eccentricity from Fig. 7, or the envelope stress factor Z from the equation in the preceding paragraph.

The deflection for the whole spring is

$$y = \frac{PD^3\pi(N - 1.2)}{4GH}$$

$$y = \frac{8PD^3(N - 1.2)}{Gd^4}$$

for round bars, where $N - 1.2$ is the number of effective turns. The deduction of 1.2 for inactive end turns is obtained from the foregoing analysis by calculation of y_a.

Discussion

H. C. PERKINS.[4] Elastic curves of helical compression springs, as computed by the author, check so well with the experimental results[2] reported by Pletta and Maher that the essential validity of his theory seems definitely established. He is to be congratulated upon the success of his analysis and the accuracy of his calculations. Nevertheless, his formulas are inconvenient, especially in case his Fig. 7 is inapplicable.

Since the author's theory seems to be of sufficient interest and value to warrant an attempt to simplify the analysis and reduce its results to more convenient form, the following discussion of his paper has three objects:

1 To check the author's theory
2 To simplify his analysis
3 To derive formulas which could be used in design.

Fig. 11 of this discussion represents the lower half of a compression spring stressed and deflected by a force P with eccentricity e and supported by reactions P_3 and P_4 at the lower end coil as shown. Since the bending moment does not, in general, vanish at the point $V = \pi$, but is small and changing rapidly, the author chose to take the origin O at the cross section of zero bending moment and to regard V as undetermined.

It therefore follows that the stress resultant at O (including any reaction between O and b) is a vertical force P_4 with action line intersecting a coil diameter through O, at an unknown distance C from the axis of the coil. If f represents the number of free turns of wire between end contacts b, then half the free length of the wire is equal to $Rf\pi$ as suggested in Fig. 11 of this discussion.

Three independent equations of equilibrium may be written. Thus, summing forces

$$P_3 + P_4 - P = 0 \quad\dots\dots\dots\dots [14]$$

Summing moments about a coil diameter through a

$$P_3 R \sin f\pi + P_4 C \sin U = 0 \quad\dots\dots\dots [15]$$

Summing moments about a coil diameter through b

$$Pe \sin f\pi - P_4 C \sin V = 0 \quad\dots\dots\dots [16]$$

But a moment sum about the coil diameter perpendicular to the one through a is also useful in case $\sin f\pi$ and $\sin V$ vanish as they do when f = integer. Thus

$$Pe + P_3 R \cos f\pi + P_4 C \cos U = 0 \quad\dots\dots [17]$$
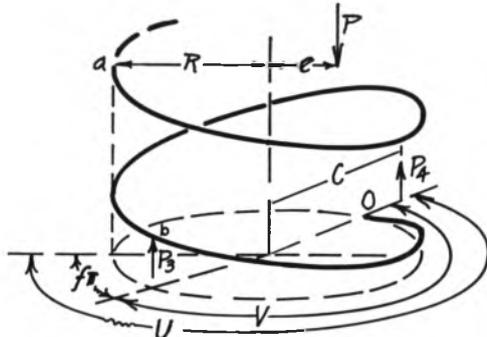
FIG. 11 LOWER HALF OF COMPRESSION SPRING STRESSED AND DEFLECTED BY FORCE P AND SUPPORTED BY REACTIONS P_3 AND P_4 AT THE LOWER END COIL

Before proceeding with the general analysis, consider two special cases.

In case 1, when $2f$ is an odd integer, the end contacts b are at opposite ends of a coil diameter. In this case, by symmetry of the upper and lower halves of the spring $e = 0$, since P must be equidistant from the two end contacts. Then by Equation [16] of this discussion, since $f\pi = 0$ and $P_4 C$ does not vanish, it follows that $\sin V = 0$, so $V = \pi$ and U = (integer) (π) + 90 deg. This is 2 deg smaller than the author's value. The error is not one of calculation but seems to follow from the assumption that the radial slope λ vanishes at the origin. This assumption is, in general, inconsistent with the fact that, as shown by the author, there are torques on most cross sections of the "dead end coil" of an actual spring, but the tests[2] of Pletta and Maher indicate that the consequent errors are probably smaller than those introduced in the manufacture of the spring.

In case 2, when f is an integer, $\sin f\pi = 0$, and since $P_4 C$ does not vanish, Equation [16] of this discussion requires that $\sin V = 0$, so $V = \pi$.

Given the value of V as in cases 1 and 2, the corresponding value of C/R is found from the condition that the deflection y must vanish at b. The deflection y at any point of the elastic curve may be formulated as follows.

The bending and twisting moments on any cross section located by angle θ from 0 or angle ϕ from b are

$$M = P_4 C \sin \theta + P_3 R \sin \phi \quad\dots\dots\dots [18]$$

$$T = P_4 (C \cos \theta - R) - P_3 R (1 - \cos \phi) \quad\dots\dots [19]$$

and corresponding values for unit downward force on the elastic curve at angle α from O or β from b are

$$m = -R \sin (\alpha - \theta)$$
$$= -R \sin (\beta - \phi) \quad\dots\dots\dots\dots\dots [20]$$

$$t = -R [1 - \cos (\alpha - \theta)]$$
$$= -R [1 - \cos (\beta - \phi)] \quad\dots\dots\dots\dots [21]$$

Substituting these values in the strain-energy formula

$$y = \int \frac{Mm\,dL}{EI} + \int \frac{Tt\,dL}{GH} \quad\dots\dots\dots [22]$$

and integrating $dL = Rd\theta = Rd\phi$ from $\theta = 0$ to $\theta = \alpha$ or from $\phi = 0$ to $\phi = \beta$, the result is

$$\begin{aligned} y = &P_4 R^3 \frac{c}{2R} \left[\frac{1}{EI} + \frac{1}{GH} \right] (\alpha \cos \alpha - \sin \alpha) \\ &+ \frac{P_4 R^3}{GH} (\alpha - \sin \alpha) + \frac{P_3 R^3}{2EI} (\beta \cos \beta - \sin \beta) \\ &+ \frac{P_3 R^3}{2GH} (2\beta - \beta \cos \beta - 3 \sin \beta) \quad\dots\dots [23] \end{aligned}$$

Note that by taking $\beta = \theta$, $P_3 = W$, $P_4 = 0$, and $\theta_0 = 0$, or taking $\alpha = \theta$, $P_4 = W$, $P_3 = 0$, $C = R$, and $\theta_0 = 0$, Equation [23] and the author's Equation [1] are reduced to the same form.

Since P_4 is the only reaction if the angle θ is less than V, the zero deflection at b is formulated by letting $P_3 = 0$ and $\alpha = V$ in Equation [23] of this discussion. It follows thus that

$$\frac{C}{R} = \frac{2EI}{EI + GH} \frac{V - \sin V}{\sin V - V \cos V} \quad\dots\dots\dots [24]$$

or if $\dfrac{GH}{EI} = 0.77$

$$\frac{C}{R} = 1.13 \frac{V - \sin V}{\sin V - V \cos V} \quad\dots\dots\dots\dots [25]$$

In cases 1 and 2, $V = \pi$ and $C/R = 1.13$. In general, however, before C/R can be evaluated, V must be computed from the condition that, by symmetry, the radial slope λ must vanish at the mid-section. Thus, bending and twisting moments on any cross section are given by Equations [18] and [19] of this discussion and corresponding values for unit torque at mid-section a are

$$m = -\sin (U - \theta) = -\sin (f\pi - \phi) \quad\dots\dots [26]$$

$$t = \cos (U - \theta) = \cos (f\pi - \phi) \quad\dots\dots\dots [27]$$

Substituting these values in the strain-energy equation

$$\lambda_a = 0 = \int \frac{Mm\,dL}{EI} + \int \frac{Tt\,dL}{GH} \quad\dots\dots [28]$$

integrating $dL = Rd\theta = Rd\phi$ from $\theta = 0$ to $\theta = U$ or from $\phi = 0$ to $\phi = f\pi$, solving for $\dfrac{-RP_3}{CP_4}$ and equating it to the corresponding ratio from Equation [15] of this discussion

$$\frac{-RP_3}{CP_4} = \frac{U \cos U - A \sin U}{f\pi \cos f\pi - \sin f\pi} = \frac{\sin U}{\sin f\pi} \quad\dots\dots [29]$$

in which $A = \dfrac{EI \left(2 \dfrac{R}{C} - 1 \right) + GH}{EI + GH}$

To check, let $\dfrac{GH}{EI} = 0.77$, and take $f = \frac{1}{2}$ which comes under

case 1. Then $\dfrac{C}{R} = 1.13$, $A = 0.87$, $\sin f\pi = 1$, $\cos f\pi = 0$, $\sin U = -1$, $\cos U = 0$. These values do not satisfy the equation. To estimate the error, regard $f\pi$ as undetermined. Then

$$\cot f\pi = \frac{0.13}{f\pi}$$

Solving by successive approximations, $f\pi = 94$ deg 30 min. The author found that $e = 0$ when $U - V = \text{(integer)} (\pi) + 92$ deg. In general, by taking $A = 1$ in Equation [29] of this discussion, the error is eliminated in case 1 and perhaps in other cases as well. If this change is effected, the equation can be written in the form

$$\cot U = \frac{f\pi \cot f\pi}{U} \quad\dots\dots\dots\dots\dots [30]$$

which is easily solved by successive approximations.

For example: let $f = 3.25$ so $\cot f\pi = 1$. Substitute $U_0 = 585 + 180$ deg on the right and solve for U_1 on the left. Thus

$$\cot U_1 = \frac{585\,(1)}{765} = 0.764$$

so $U_1 = 772.6$. Substitute U_1 on the right and find $U_2 = 772.9$ deg on the left. This completes the solution, since further refinement is useless for slide-rule computation. Now $V = 772.9 - 585 = 187.9$ deg, which seems to check with the corresponding ordinate in the author's Fig. 7.

Using this value of V in Equation [25] of this discussion, $C/R = 1.238$

Equations [14] and [15] of this discussion may now be solved to find $P_4 = 0.583P$ and $P_3 = 0.417P$ and Equation [16] solved to find $e/R = 0.101$ which is about 10 per cent less than the corresponding ordinate in the author's Fig. 7. Finally, the over-all deflection of the spring is obtained by doubling the deflection at mid-section, the latter being found from Equation [23] by taking $\alpha = U$ and $\beta = f\pi$.

A formula approximating the author's "envelope curve" may be found by taking $f =$ an integer and solving Equations [24], [14], [17], and [29] to find

$$\frac{e}{R} = \frac{-2}{f\left(3 + \dfrac{GH}{EI}\right) + 2} \quad\dots\dots\dots\dots [31]$$

Disregarding sign and taking $\dfrac{GH}{EI} = 0.77$, the results check the author's values of e/R for $f = 1, 2$, and 3 free turns, respectively. Increasing the numerator 5 per cent and, changing sign, the "envelope formula" is

$$\frac{e}{R} = \frac{2.1}{f\left(3 + \dfrac{GH}{EI}\right) + 2} \quad\dots\dots\dots\dots [32]$$

Since e, in Fig. 11 of this discussion, is alternately positive and negative as f is alternately odd and even, and passes through zero when $2f$ is an integer, the curve e versus $f\pi$ is periodic. It may be approximated by the formula

$$\frac{e}{R} = \frac{-2\cos f\pi}{F\left(3 + \dfrac{GH}{EI}\right) + 2} \quad\dots\dots\dots\dots [33]$$

in which F is the integer nearest to the given value of f.

In the foregoing example, in which $f = 3.25$ and $\dfrac{GH}{EI} = 0.77$, $F = 3$ and the approximate eccentricity ratio is

$$\frac{e}{R} = \frac{1.414}{3 \times 3.77 + 2} = 1.06$$

This discussion seems to support the conclusion that, although simplifying assumptions have introduced some errors, the errors are probably small. The author's calculations also seem to be highly accurate despite their complication. However, as shown, the analysis need not be complicated and, judging by the illustrative example near the close of this discussion, the formulas therein derived could easily be used in the design of helical compression springs.

Author's Closure

Professor Perkins' discussion is of much interest as showing a different method of approaching the problem. His treatment is admittedly shorter than the author's, and possibly simpler, but it has been compressed to such an extent that it is more difficult to follow. For instance, Equation [23] is given for the elastic curve, but no equations expressing P_3 and P_4 directly in terms of P are given. Apparently it would be necessary to use Equation [15] which involves U, finding U from Equation [30]. The author admits frankly that the assumption of zero λ at the origin is not exactly true. However, it is very nearly true, physically, as is the case with many assumptions commonly accepted in elasticity calculations. It is not clear why the reaction P_4 has been taken as acting outside the coil.

The paper is concerned primarily with the determination of the elastic curve, as the subject indicates. No attempt was made to give working formulas for practical spring calculation. For such purposes, neither the author's paper nor Professor Perkins' discussion is suitable. A further simplification is necessary and the results must be put in tabular form so that the use of formulas is eliminated. Such tables have been prepared by the author and used for several years in routine spring calculation. They have proved to be rapid and accurate, containing the proper allowance for inactive end turns and load-eccentricity effect, as expressed in the envelope curve.

Helix Warping in Helical Compression Springs

By D. H. PLETTA[1] and F. J. MAHER,[2] BLACKSBURG, VA.

When helical springs are compressed between parallel planes, the load is usually eccentric with the longitudinal axis and the helix, in its deflected position, is no longer truly helical but assumes a sinusoidal or warped shape. The magnitude of the eccentricity and of the warping is a function of the number of turns in the spring.

The research reported on here illustrates this warping for a heavy helical spring, whose length was varied for the series of tests, and it also compares the test results with the theoretical behavior as outlined in the accompanying paper of H. C. Keysor. In general, the actual and theoretical behavior were in excellent agreement, thus apparently substantiating the theory. An analysis of the data also emphasizes the need of always assuming some eccentricity of load in practical design, to allow for the lack of perfect concentricity of the individual turns.

THE TESTS reported on in this paper were undertaken to determine the correlation between the actual behavior of loaded helical compression springs and the theoretical behavior as outlined in the companion paper[3] by H. C. Keysor. The theory indicates that helical springs, when compressed between parallel planes, are usually subjected to an eccentric rather than a concentric load, and that this eccentricity causes the helix to warp under load. The term warping, as used here, indicates that the deflected helix assumes a sinusoidal shape and does not remain truly helical in its deflected position as the ordinary spring formula leads one to believe. Fig. 1 illustrates this phenomenon. Curve A shows the unloaded helix "unwound" and projected on a vertical plane. Curve B indicates the deflected position of this helix after application of the load, the solid curve showing actual behavior, and the dotted line showing the commonly assumed position. The shaded area between curves A and B represents deflection and is replotted in curve C for only half of the helix.

The actual and theoretical behavior of compression springs may then be checked by measuring the actual eccentricity of load or by measuring the warping of the helix, and comparing these observations with the theoretical values. The tests as carried out included only observation of helix warping, mainly because apparatus capable of determining warping already existed and because it would have been somewhat difficult to measure eccentricity accurately on the heavy stiff spring used.

A single spring was used in this project. Hence, all variables such as differences in heat-treatment, moduli of elasticity, coil-and-bar diameter, etc., common to any series of separate speci-

mens were eliminated or at least minimized. The bar as coiled consisted originally of about seven turns from each end of which equal lengths were cut as the spring was shortened for the series of tests. The spring had an outer diameter of 6.42 in. and a bar diameter of 1.336 in. A plain carbon steel (0.9 to 1.05 per cent C) was used, coiled hot on a slightly tapered mandrel, and subjected to an oil quench from 1525 F and a subsequent drawing to 850 F.

The spring lengths, as tested, were placed between special helical end blocks mounted to the base and movable head of a testing machine, Fig. 2. Very thin plaster of Paris cappings, as cast in place under load, insured continuous contact for the first 270 deg of the dead turn. The balance of the apparatus to measure the warping consisted essentially of a collar, concentric with and free to rotate about the lower end block, to which a vertical post was attached. An Ames dial reading to 0.001 in. was fastened to a sleeve which was free to slide up and down the post. A tapered pin, which fits through the sleeve and post simultaneously, could be inserted in any one of a series of holes in the post spaced 1 in. apart. The cantilever arm attached to the lower end of the Ames-dial plunger rested on the top side of
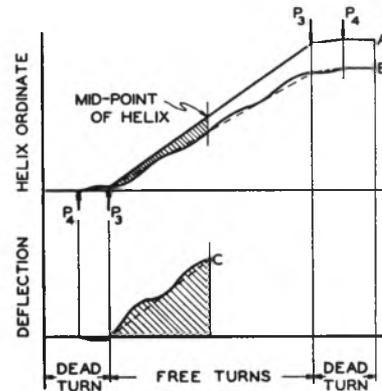
FIG. 1 SINUSOIDAL SHAPE ASSUMED BY DEFLECTED HELIX

the wire of the spring. With this arrangement, it was possible to determine the elevation of any point on the coil to 0.001 in. above a datum for any given load. A subtraction of two successive elevations at any given point gave the net deflection for that increment of load.

During a test run, readings were taken at points spaced 30 deg apart along the helix for loads of 500 and 8000 lb. The cantilever arm was used to eliminate the error that could possibly result if points on the outside of the wire had been used. Any twist in the wire, when the coil was loaded, would have caused these points to deflect more or less than the helical axis, whereas any measurement made on a horizontal tangent to the top of the wire would always remain half a wire diameter above the axis. At each load the deflection of the whole spring was also observed by noting the movement of the head of the testing machine.

At the 8000-lb load, the upper and lower split wedge blocks were adjusted so that the deflection of the end of the dead

---

FIG. 2 METHOD OF MOUNTING SPRING LENGTHS IN TESTING
MACHINE

with the exception of case 2, were such that the eccentricity was the maximum attainable. The theoretical spring deflection was calculated, using a shearing modulus of elasticity of 11,600,000 lb per sq in.

Fig. 4 shows the helix warping up to mid-length of the spring for the seven tests outlined in Table 1. The other half of the spring should, of course, be identical and tests indicated that it was, but there seemed little need of plotting the curves for the entire spring. The load increment was 7500 lb. The solid line represents theory. No attempt was made to draw a curve through the data points. In four of the tests the correlation between theoretical and actual behavior was almost perfect, whereas in the other three cases it was only fair. There are
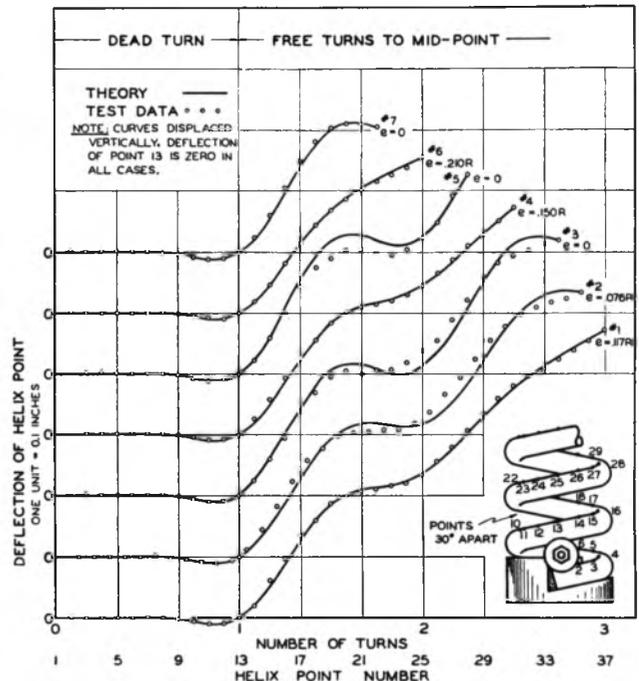
FIG. 4 HELIX WARPING UP TO MID-LENGTH OF SPRING

turn (point 13) was reduced to zero relative to its seat. The wedges can be seen in Fig. 2 and are also illustrated in detail in Fig. 3. They resemble a bolt with the circular head of the bolt and washer tapered. Under the 8000-lb load the two ends of the dead turn closed several hundredths of an inch allowing point 13 to deflect. This movement was reduced to zero by tightening the nut, thus separating the ends of this turn. Various other nonadjustable supports were tried but all of them allowed some deflection due to the localized contact deformation. It was thought best to have this deflection zero in keeping with the theory, but it should be added that the small movement allowed by these other supports was not critical, for the deflections of points along the helix were only slightly affected.

FIG. 3 SPLIT WEDGE BLOCK

Table 1 indicates the scope and some of the results of the tests. Seven tests were made, the bar length varying from 6 to 3.511 total turns. Three of these lengths were such that the eccentricity of the load was theoretically zero; whereas the others,

several reasons why the agreement in these last three cases should not be perfect. The measurement of the deflection on the top of the bar rather than along the neutral axis; the slight rocking of the bar on the split wedge blocks, tending to shorten the spring; the imperfect seating on the helical end seats; and the lack of perfect concentricity for the individual turns of the spring, all tended to modify perfect behavior. The last condition is the most important. It is practically impossible to coil heavy bar so that all turns are identical in diameter, pitch, and concentricity. If only one turn is displaced laterally, an axial load on the spring would be eccentric on this coil. Theoretically the warping decreases near the center of the spring as the ec-

TABLE 1 SCOPE AND RESULTS OF COMPRESSION TESTS ON HELICAL
SPRINGS

Test no.	Number of turns			Deflection in in.		Variation from theoretical, per cent	Theoretical eccentricity
	Free[a]	Active (theoretical)	Excess	Actual	Theoretical		
1	4	4.415	0.415	0.947	0.938	+0.96	0.1171R[b]
2	3.75	4.194	0.444	0.888	0.892	−0.45	0.0763R
3	3.511	3.979	0.468	0.841	0.846	−0.59	0
4	3	3.399	0.399	0.732	0.723	+1.39	0.1503R
5	2.511	2.979	0.468	0.625	0.634	−1.42	0
6	2	2.371	0.371	0.507	0.504	+0.59	0.2096R
7	1.511	1.979	0.468	0.416	0.421	−1.19	0

[a] All test springs had two dead coils in addition.
[b] R = mean coil radius.

centricity of load increases. Hence the sinusoidal effect would tend to be damped out especially at the center of the spring. The measured eccentricity of individual turns on the spring amounted to several hundredths of an inch, which corresponded approximately to $e = 0.02R$ rather than zero. Naturally the magnitude of this error was relatively larger in cases 3 and 5 when the eccentricity should have been zero than in cases 1, 4, and 6 when e was appreciable. Yet in case 7, with $e = 0$ theoretically, the correlation is very good. The reason for this is that this spring had only 1.511 free turns and was therefore too short to be affected by any lateral displacement of a full turn, the short length practically requiring concentricity with itself.
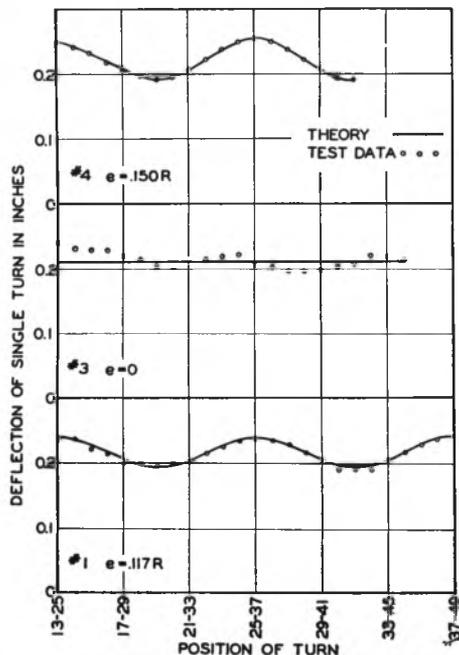
FIG. 5   DEFLECTION OF INDIVIDUAL TURNS 30 DEG APART

The actual eccentricity of individual turns in cases 3 and 5 was actually determined, but its existence would also be suspected when studying Fig. 5. Here the deflection of individual full turns 30 deg apart is plotted for cases 1, 3, and 4. The

sinusoidal nature of the test data for case 3 is possible only when the load is eccentric. When considering this variation, it should be remembered that the total load on the spring, considered as a concentrated force, acts perpendicular to the diameter through the mid-point of the spring, and that it is generally displaced from the longitudinal axis of the spring.

Some objection to these tests and the theory might be raised
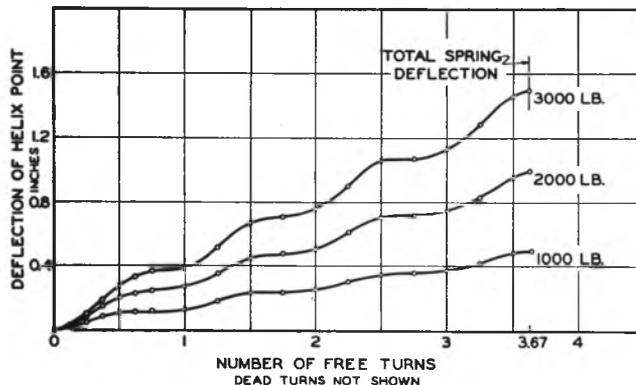
FIG. 6   HELIX WARPING OF A COMMERCIAL SPRING

because of the difference in the end conditions between these tests and those in commercial springs with their tapered ends ground square. Fig. 6 illustrates helix warping in a commercial spring tested and reported on previously.[4] Here again the sinusoidal effect is quite evident.

It is the belief of the authors that the tests reported in this paper substantiate Mr. Keysor's theory on spring behavior and also emphasize the absolute necessity of always assuming some eccentricity for the load in commercial design to be on the safe side. In other words, an eccentricity based on the envelope of the theoretical cyclic variation in eccentricity,[5] rather than the theoretical eccentricity itself, should be used to allow for the lack of concentricity arising in the practical manufacture of springs. In the commercial design it should, of course, be remembered that eccentric loads produce greater shearing stress than the commonly assumed axial loads.

---

[4] "The Effect of Overstrain on Closely Coiled Helical Springs and the Variation in the Active Coils With Load," by D. H. Pletta, S. C. Smith, and N. W. Harrison, Bulletin 24, Engineering Experiment Station, Virginia Polytechnic Institute.

[5] Reference 3, Fig. 7.

Boundary Film Investigations

By SYDNEY J. NEEDS,[1] PHILADELPHIA, PA.

This paper describes attempts made in Dr. Kingsbury's laboratory to detect the influence of boundary surfaces on the viscosity of thin films of lubricant between two optically plane parallel disks. Investigations included both mineral and vegetable oils. Details of two testing machines are given and typical results with each machine are described.

INTRODUCTION

THE MANNER in which a liquid in contact with a solid surface is influenced by that surface has been the subject of considerable study in this country and abroad. It appears obvious that the molecules of the liquid actually in contact with and adhering to the surface, or to adsorbed layers on the surface, must necessarily behave differently from adjacent molecules of the liquid in bulk. But the nature of this surface phenomenon and the distance from the surface to which its influence extends are not well understood and experimental studies have given rise to widely divergent opinions on these particular points. Kingsbury (1)[2] pointed out some 37 years ago that the nature of this surface property was probably "an intensified viscosity in that part of the fluid within the region of attraction of the surface molecules of the metal." Hardy and Nottage (2) offer proof "that the attraction field of the solids modify the state of the lubricant throughout a layer many hundreds or thousands of molecules in thickness;" and they place the limit of boundary lubrication (3, 4) at a distance of 0.007 to 0.010 mm (0.0003 to 0.0004 in.), that is, the film must be of that order of thickness if any of the liquid is to be beyond the range of surface influence. Wilson and Barnard (5) found that capillaries as large as 0.3 mm (0.01 in.) diam would gradually clog up under the flow of oils and would close completely if a small percentage of oleic or stearic acid were added to the oil. This clogging was attributed to the gradual building up of adsorbed films on the walls of the capillaries.

In a carefully conducted series of experiments with capillaries of approximately the same diameter as those used by Wilson and Barnard, Bulkley (6) was able to show that the clogging was due to the fact that the oil had not been thoroughly filtered. From further experiments with highly filtered oils in fine glass and platinum capillaries having diameters of the order of 0.01 mm to 0.03 mm, Bulkley concluded "that the thickness of any fixed layer adjacent to the walls is not in excess of two or three hundredths of a micron, or approximately a millionth of an inch." Confirming results have been published by Bastow and Bowden (7), who conclude from measurements of the flow of liquids between optically polished flat plates that no sign of induced rigidity is to be detected in liquids at a distance of 1000 Å [4(10)$^{-6}$ in.] from the surface. There are many other reports in the literature which form, as pointed out by Bowden (8), "a large body of conflicting evidence and opinion as to the range of surface forces."

The present paper reports some of the results of a 4-year study of thin films of vegetable and mineral oils conducted in Dr. Kingsbury's laboratory under his direction. Aside from the problems of filtration, selection of the proper metal for the test surfaces and the preparation of the surfaces, the experimental work divided itself into two distinct parts. Part I of this paper describes attempts to discover indications of increasing viscosity in thin films of oil formed by two plane disks approaching each other under loads of the order of 0.20 lb per sq in. with no tangential motion. Part II describes experiments with plane surfaces approaching under loads up to 800 lb per sq in. with continuous relative rotation of the surfaces. It also describes a method of measuring the minimum film thickness reached under these conditions. Experiments reported in part II are extremely simple and may be readily reproduced. They add clarity to the results of part I and appear to show that steel surfaces have a profound influence on the behavior of lubrication films up to thicknesses greater than 60 millionths of an inch (0.0015 mm); the thickness depending somewhat upon the lubricant. These films have great mechanical strength and are capable of supporting loads at least as great as 800 lb per sq in. Under such loads, the film thickness quickly reaches a constant minimum value which remains unchanged indefinitely even with continuous relative rotation of the surfaces.

The influence of the surfaces on the film manifests itself in what might be termed "directional rigidity," a property, which to the author's knowledge, has not been previously observed. As mentioned, the condition of constant minimum film thickness is indicated by constant torque at a given rpm, provided the rpm is above some minimum value. As long as the film is under continuous shear at a rate above this minimum, it behaves as a purely viscous liquid in the circumferential direction and the torque is proportional to the rpm. In the radial direction, however, the liquid exhibits remarkable plasticity or rigidity since, under heavy loads, it will not flow radially from between the surfaces. Building up of an apparently similar rigidity in the circumferential direction may be observed by stopping rotation and running at infrequent intervals only long enough for a torque observation to be made. The torque will increase rapidly at first and then more slowly, finally reaching an upper limit within 1 to 3 days. This upper limit will be several times greater than the constant torque when the film is free from circumferential rigidity. During the rigid stage, the film will resist considerable circumferential stress without appreciable deformation. If continuous rotation is again resumed, the circumferential rigidity will be gradually removed; the torque will decrease rapidly at first and then more slowly until finally the same constant value is reached when rigidity has entirely disappeared, thus indicating no change in film thickness. In any particular test this cycle may be repeated as often as desired.

Oils used in the earlier tests were rotated for periods up to 50 hr in a clinical centrifuge. A small ultracentrifuge, operating at extremely high speeds, was next constructed and used. In the later tests the samples of oil were removed by pipette from near the centers of containers that had remained undisturbed for several months. It may therefore be assumed that the oils were used in the same condition as commerically produced. Centrifuging of the oils made no apparent difference in test results.

Theoretical

Formulas for two plane, circular, parallel plates, approaching each other under the action of an external force and with the space between the plates filled with a viscous liquid, may be readily developed from Reynolds' (9) basic equation

$$\frac{\partial}{\partial x}\left(h^3 \frac{\partial p}{\partial x}\right) + \frac{\partial}{\partial z}\left(h^3 \frac{\partial p}{\partial z}\right) = 6\mu \left\{(U_0 - U_1)\frac{\partial h}{\partial x} + 2V_1\right\} \dots [1]$$

Geometry of the surfaces is shown in Fig. 1. The notation used is as follows:

a = radius of plane circular plates
f = shear stress in liquid at surfaces of plates due to relative rotation
f_r = shear stress in liquid at surfaces of plates due to radial pressure flow
h = film thickness or uniform distance between parallel plates
N = revolutions per unit time of rotating plate
p = pressure at any point s in film above atmospheric or uniform ambient pressure
r = radial distance from centers of surfaces to point s. Rectangular coordinates of s are x, z.
T = torque due to relative rotation of plates
t = time required for any given decrease in film thickness
U = $2\pi r N$ = linear velocity at s due to relative rotation of surfaces
$U_0 - U_1$ = relative tangential velocity of surfaces
V_1 = relative normal velocity of surfaces
W = external load acting on surfaces
μ = viscosity of the liquid between the surfaces; assumed constant and uniform.
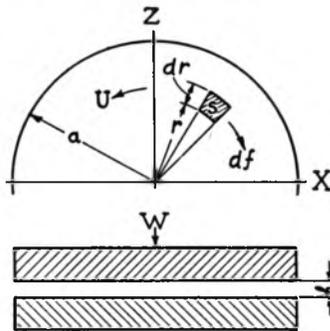
FIG. 1 MATHEMATICAL NOTATION; PLANE CIRCULAR PLATES

Since there is no relative tangential velocity of the surfaces in the case under consideration, $U_0 - U_1 = 0$; and, since the plates are parallel, there is no variation of film thickness; hence, $\partial h/\partial x = 0$. By definition $V_1 = dh/dt$; hence Equation [1] reduces to:

$$\frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial z^2} = \frac{12\mu}{h^3}\frac{dh}{dt}$$

a solution of which is

$$p = \frac{3\mu}{h^3}\frac{dh}{dt}(x^2 + z^2 - a^2) \dots \dots [2]$$

The pressures are symmetrical with respect to the centers of the plates; hence, the total load W will be four times the integrated pressure over each quarter. Therefore

$$W = \frac{12\mu}{h^3}\frac{dh}{dt}\int_0^a \int_0^{\sqrt{a^2-x^2}} (x^2 + z^2 - a^2)\, dx\, dz \dots [3]$$

Successive integration gives

$$W = -\frac{3\pi\mu a^4}{2h^3}\frac{dh}{dt} \dots \dots [4]$$

and

$$t = \frac{3\pi\mu a^4}{4W}\left(\frac{1}{h_2^2} - \frac{1}{h_1^2}\right) \dots \dots [5]$$

the time required for the surfaces to fall from h_1 to h_2. If h_2 is small compared with h_1, the initial film thickness may be neglected and Equation [5] becomes

$$t = \frac{3\pi\mu a^4}{4Wh^2} \dots \dots [6]$$

from which

$$h = \frac{a^2}{2}\sqrt{\frac{3\pi\mu}{W}}\sqrt{\frac{1}{t}} \dots \dots [7]$$

Equations [5] and [6] were first developed by Stefan (10) in his studies of the approach of plane, circular disks separated by a viscous liquid. Reynolds' Equation [1], published 12 years later is of great general utility since, in addition to the present case, it is also the basis of modern theoretical treatment of plane and cylindrical bearings.[3]

From (9) the shear stress at the surfaces of the plates, due to outward radial flow of the liquid, is

$$f_r = \frac{h}{2}\frac{dp}{dr} \dots \dots [8]$$

and this stress will be the same on any radius passing through the common axis of the plates. Since $x^2 + z^2 = r^2$, we have from Equation [2]

$$\frac{dp}{dr} = \frac{6\mu}{h^3}\frac{dh}{dt}r$$

and from Equation [4]

$$\frac{dh}{dt} = -\frac{2h^3 W}{3\pi\mu a^4}$$

Substituting in [8]

$$f_r = -\frac{2hWr}{\pi a^4} \dots \dots [9]$$

the negative sign indicating that the shear stress opposes flow from between the surfaces. Since the shear stress is maximum when $r = a$

$$f_{r(\max)} = -\frac{2hW}{\pi a^3} \dots \dots [10]$$

The film thickness may also be calculated by rotating one of the plates relative to the other at a known rate and measuring the required torque.

Referring to Fig. 1, the frictional drag opposing rotation at the differential area $2\pi r dr$ is df, and the moment is $r df$. If N is expressed in revolutions per minute, the linear velocity U is $2\pi r N/60$. Then for a simple Newtonian liquid

$$df = \frac{2\pi\mu U r dr}{h}$$

---

[3] Reynolds' original paper contains many misprints, which unfortunately also appear in his "Collected Papers." His expression, similar to Equation [5], for the approach of parallel elliptical plates should have the coefficient 3/2 instead of 3.

and

$$rdf = \frac{\pi^2 \mu N}{15h} r^3 dr$$

The total torque

$$T = \frac{\pi^2 \mu N}{15h} \int_0^a r^3 dr$$

from which

$$h = \frac{\pi^2 \mu N a^4}{60T} \dots \dots \dots \dots \dots \dots [11]$$

At any time during a test Equation [11] may be used as a check on Equation [7].

EXPERIMENTAL

I—THE DISK VISCOMETER

The disk viscometer, shown in Fig. 2, was constructed to measure the rate of approach of two plane disks under conditions that would enable a comparison to be made with theory. It consists essentially of two disks A with plane surfaces and the balanced beam B, with a micrometer screw C to measure the movement of a pin fastened in the end of the beam. The upper plane is $7/8$ in. diam and is firmly attached to the upper bar D of the rugged steel frame by means of a fitting electrically insulated from the frame. The lower plane is $3/4$ in. diam and has a steel ball placed

SCALE – INCHES

FIG. 2   DISK VISCOMETER

so that the lowest point on its surface is located as near as possible to the center of gravity of the disk. The ball rests on a plane support attached to the beam, thus locating the disk and permitting its plane surface to adjust itself to the upper plane. A thin-steel cup E is provided to prevent loss of liquid flowing from between the approaching surfaces. The beam is supported by the knife-edge F. Load is applied by the $1/10$-lb sliding weight G and may be varied by changing the position of the weight. When the weight is against the knife-edge fulcrum H, the center of gravity of the beam and all its parts is brought as close to the knife-edge as possible by adjustments of the counterweights J. The cam K is used to raise the loaded end of the beam, thus keeping the surfaces separated until the start of an experiment.

The micrometer-screw nut is fastened to the upper bar by four screws and electrically insulated with sheet mica. A lower nut is provided to eliminate lost motion in the screw. The micrometer screw is $1/2$ in. diam and carefully cut with 40 threads per in. Smooth operation is obtained by lapping the screw and nuts together. After assembly, the spindle in the end of the micrometer screw was lapped optically flat. A disk L, $5^1/2$ in. diam and graduated into 25 equal numbered divisions, is attached to the upper end of the micrometer screw. Each division is subdivided into tenths and a vernier M is provided for reading to $1/10$ of each subdivision or $1/100,000$-in. vertical movement of the screw. The distance from the micrometer screw to the knife-edge is 10 times greater than the distance from the knife-edge to the center line of the plane disks. Each vernier division therefore represents a change in film thickness of one millionth of an inch. Contact between micrometer and beam is indicated by a milliammeter in a circuit closed when the micrometer spindle touches the pin in the beam. A similar circuit is provided to indicate contact between the plane surfaces. By loosening the frame nut N, the upper bar may be swung to one side permitting insertion of the lower disk or removal of the beam.

The viscometer is operated in a wooden box (not shown) covered with aluminum foil and heavy paper for heat insulation. Curtained inspection windows are fitted in the box and the necessary controls are provided for external operation of the instrument. The box is electrically heated by coils operated through a relay and a differential-expansion thermostat capable of holding the temperature constant to within 0.1 F. Viscometer temperatures are measured by thermometers O located in mercury wells in the vertical supports at each end of the instrument and also by a thermometer suspended in the atmosphere near the plane surfaces.

The plane disks were first made of normalized low-carbon steel and later a second pair was made of hardened tool steel. The surfaces were optically polished and flat to better than one millionth of an inch, including the edges. Turned edges were avoided by cementing the otherwise finished disk with beeswax in a ring about $2^1/4$ in. diam made from the same bar of steel as the disk, and finishing the two as a single piece. The turned edge will then be on the ring and, if the diametral clearance between the disk and ring is of the order of 0.001 in., the plane surface of the disk will be continuous to its edge. After cementing the disks in the rings, the surfaces were ground flat, lapped on lead, and finally polished with optical rouge on pitch. Supports were then placed under the unpolished sides of the disks and the rings gently heated until melting of the beeswax permitted them to drop free, leaving the edges of the disks untouched.

Most of the difficulties encountered in polishing the surfaces to the required flatness proved to be simply questions of technique; eventually the point was reached where the scratching of a pair of surfaces, instead of actually threatening the success of the investigation, occasioned merely a relatively short interruption in the experimental program. However, researches for a
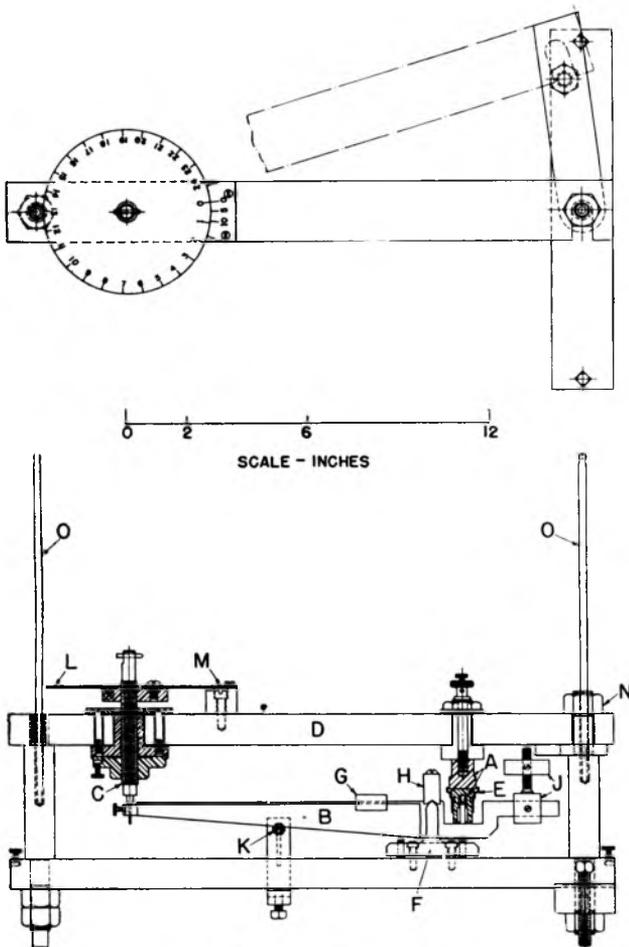
satisfactory metal of which to make the disks were not so successful. There appears to be no metal which, when polished and examined even under low powers of the microscope, will present a surface of uniform appearance even remotely approaching that of well-polished glass. The metal surfaces invariably contained pits or nonmetallic inclusions, iron carbides or traces of manganese in the form of gray spots with well-defined but irregular edges. At times traces of corrosion were also visible somewhat resembling nonmetallic inclusions. No one steel of good enough grade to be considered for optical flats is likely to contain all the above flaws, but, by the very nature of the constituents of steel and its method of production, nonhomogeneous structures are to be expected on any plane. Experience finally seemed to indicate that the best surfaces were produced by high-speed tool steel or a high-chrome tool steel. The chrome steel is corrosion-resistant and will hold its shape over long periods of time but it warps badly during the hardening process and is somewhat difficult to polish. High-speed tool steel gave the most uniform surface of all metals investigated, and it was therefore used in the later experiments.

FIG. 3   APPROACH OF PLANE CIRCULAR PLATES; TESTS WITH CASTOR OIL

(Plates, $^3/_4$ in. diam; oil temperature, 92.2 F; $\mu = 50.8(10)^{-6}$ lb sec per sq in.; load, 0.199 lb. The four tests were made under identical conditions.)

After polishing, the disks were prepared by washing with pads of absorbent cotton moistened with carbon tetrachloride. When all visible traces of contamination had been removed the disks were gently rubbed with dry absorbent cotton and finally touched with a fine camel's hair brush to remove any tiny pieces of lint left by the cotton or other contamination deposited from the air. The disks were then placed in the viscometer and the upper bar locked in position. Two carefully cleaned platinum wires approximately 0.001 in. diam were placed parallel across the lower plane approximately equidistant from its center. The beam was gently lowered until the surfaces were separated only by the wires. The box was then closed, outside operating attachments were connected, and the heat turned on.

When the temperature had reached the desired constant value, a micrometer reading was taken with the wires in place. If the diameter of the wires is known, it is then a simple matter to compute the micrometer zero, which is the micrometer reading when the disks are in actual contact. This seems to be the only practicable method of determining the zero reading. If the surfaces were brought together without the wires, a false reading

would probably result due to possible surface contamination or to adsorbed air or moisture films on the surfaces. Attempts to obtain optical contact of the surfaces by wringing them together would doubtless result in scratches and would also introduce the problem of separation after the zero reading had been made. It is probable that the small-diameter wires cut through any adsorbed layers on the surfaces and came into actual contact with the metal. Assuming this to happen, it then became necessary to measure the diameters of the wires in order to establish

FIG. 4   APPROACH OF PLANE CIRCULAR PLATES;   TESTS WITH OLIVE OIL

(Plates, $^3/_4$ in. diam; oil temperature, 92.2 F; $\mu = 7.02(10)^{-6}$ lb sec per sq in.; load, 0.199 lb. The four tests were made under identical conditions.)
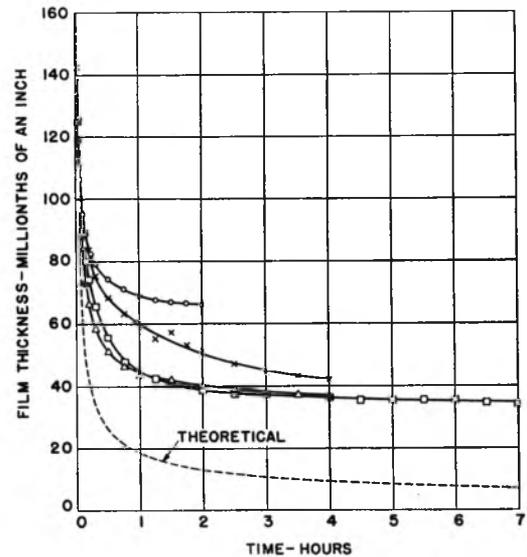
FIG. 5   APPROACH OF PLANE CIRCULAR PLATES;   TEST WITH DTE LIGHT MINERAL OIL

(Plates, $^3/_4$ in. diam; oil temperature, 92.3 F; $\mu = 5.23(10)^{-6}$ lb sec per sq in.; load 0.199 lb. The four tests were made under identical conditions.)

the micrometer zero. This was done by a filar micrometer used with a microscope. Many measurements were made at various magnifications with ordinary and monochromatic illumination and the wire diameter determined from these measurements. Assuming perfect microscopy, however, the limitations of the microscope would leave the final result open to an uncertainty of the

order of $1/2$ wave length of light and this is considerably greater than the film thicknesses it had been hoped to investigate.

After noting the zero reading, the micrometer screw was raised so that the beam could be lifted by the cam. The wires were then removed and oil introduced between the separated surfaces from a pipette. This was done with but slight temperature disturbance through a hole in the top of the box. When the temperature had again reached the constant value, the time was noted, the cam moved from beneath the beam, and the surfaces began to approach.

Observed rates of approach of the planes are compared with theory in Fig. 3 for the case of castor oil, and for olive oil and DTE light in Figs. 4 and 5, respectively. Calculations of rates of fall are based on the known bulk viscosities of the three oils and are shown by the broken lines. In general, the observed rates of approach are a fair approximation to theory but excepting the experiments with castor oil the film thicknesses at the end of 1 to 2 hr are found to be considerably greater than the calculated values. Satisfactory duplication of tests was found impossible, even though the lubricants were carefully centrifuged before using. After 8 hr, little if any reduction was observed in film thickness and, at the end of 24 hr, the film thickness was invariably found to have increased slightly above the reading at 8 hr. It is interesting to note that after a few hours the measured film thicknesses were of the same order regardless of the differences in bulk viscosities of the oils. Apparently the films were not behaving strictly according to theory. The probable causes of these discrepancies will appear from the results described under part II.

It was also observed that, when the films reached a thickness of approximately $(10)^{-4}$ in., they began to show indications of carrying current under a voltage of 0.025 volts. At thicknesses of about 60 $(10)^{-6}$ in., the resistance of the film had dropped to the point where it carried small currents as readily as the copper wire of the circuit, that is, the ammeter showed the same current flowing with and without the film in the circuit.

With loads of about 0.2 lb per sq in., the final film thicknesses were of the order of 30 to 40 millionths of an inch, subject to errors of approximately 10 millionths of an inch. Because of this high percentage of uncertainty, these figures are not entirely satisfactory. Nevertheless the above experiments indicate that there is some limiting film thickness reached by plane surfaces approaching under load. The limits of boundary lubrication placed by Hardy are considerably greater than those observed in the foregoing experiments.

II—CONTINUOUS ROTATION

It is obvious from the results cited that loads much greater than a fraction of a lb per sq in. are required to bring the surfaces to within a few millionths of an inch of actual contact. The apparatus shown in Fig. 6 was therefore constructed and mounted in the same drill press used with the tapered-plug viscometer and the journal-bearing testing machine (11, 12). The lower disk A is located by the fitting B, and held against rotation by a set screw. This assembly is located by the ball bearing C and supported by the torsion rod D. Angular twist of D, measured by a graduated curved scale E, gives an accurate measure of the torque, due to rotation of the upper disk A'. The torsion rod is supported against buckling by a split bushing F set in a ball bearing to avoid binding at the support. The torque-indicating arm is counterbalanced by the weight G and vibrations of the arm are damped by a pan of oil H. At extremely low speeds, the movement of the indicating arm may be very small; hence a microscope has been found convenient for detecting minute deflections. The upper disk is held central and rotated by the sleeve J, the bore of which has been lapped to a close sliding fit on the spindle K. Load is applied by compression of the spring L which is held
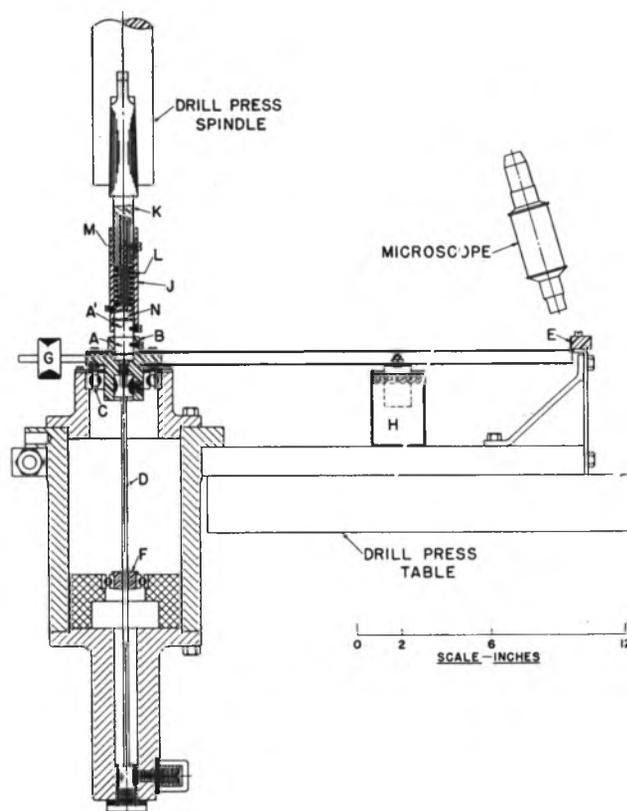
FIG. 6 TORSIONAL DISK VISCOMETER

central by the plunger M. The lower surface of the plunger is hardened and ground flat to form a contact surface for the hardened conical point of the disk support N, thus providing for adjustment of the upper disk. Both plane disks are supported on annular rings, turned on the backs of the supports. Assuming a parabolic pressure distribution in the oil film, the diameter of the ring is such as to provide support at a point directly under the center of gravity of each radial half section of the parabola. The spindle K is fitted into the drill-press spindle in the usual manner and electrically insulated by paper.

The drill press is driven by a $7^{1}/_{2}$-hp, 1200-rpm synchronous motor through a chain-and-change-gear transmission. Friction clutches are provided for starting, stopping, and reversing. An independent drive for low speeds is furnished by a small synchronous motor through a speed reducer. Speeds as low as 0.0016 rpm are thus obtained. The loading springs were calibrated in position and the spindle graduated at the top of the sleeve. A spring is provided to cover the range up to 100 lb per sq in.; a second spring is used for loads from 100 to 400 lb per sq in.; and the heaviest spring possible with the present apparatus gives loads up to 800 lb per sq in.

The hardened plane disks, $7/_8$-in. diam, were prepared as previously described, except that the final polish was made on lead instead of pitch. After cleaning, they were covered with a heavy film of oil, placed in the machine, and brought together until the oil formed a single film approximately $1/_{16}$ in. thick. At zero time, rotation was begun and load applied. Almost immediately most of the oil was pressed from between the surfaces, forming a thick ring around the edge of the film as shown in Fig. 7; thus protecting the film from external contamination. The meniscus of this ring introduced a slight negative pressure at the edge of the film.

Assuming that the plane disks remain coaxial, it may be assumed that the rate of approach is independent of relative rota-

tion. A test was therefore made with olive oil at a load of 100 lb per sq in. with the upper surface continuously rotating at 11.06 rpm. Results of the test are shown in Fig. 8. The broken line represents torque calculated on the basis that the planes are approaching at the rate indicated by Equation [7], and the circled
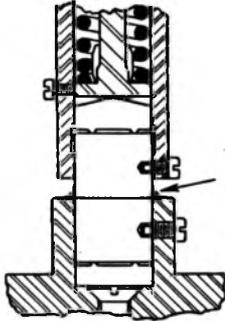
FIG. 7   OIL MENISCUS AT EDGES OF PLANE DISKS; TORSIONAL DISK VISCOMETER
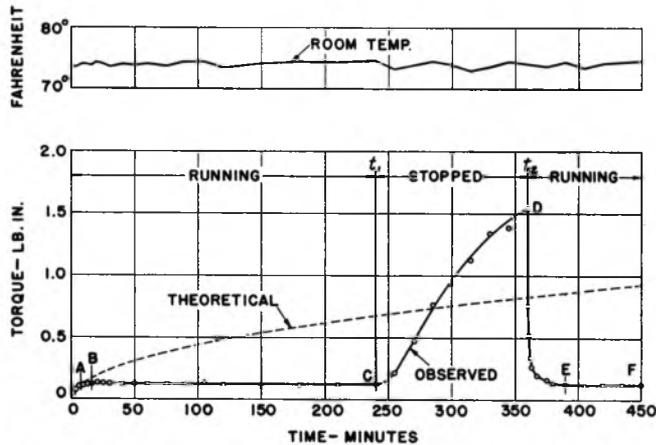
FIG. 8   TEST WITH OLIVE OIL; TORSIONAL DISK VISCOMETER
(Plane disks, $^7/_8$ in. diam; 11.06 rpm; load, 100 lb per sq in.; average temperature, 77 F; $\mu = 10.1(10)^{-6}$ lb sec per sq in.)

points are actual torque observations. During the first 8 min, over the range *OA*, the observed torque agrees with theory. This indicates that, while the film is relatively thick, the rate of approach is the same as calculated. It is interesting to note, however, that after 16 min the torque becomes constant over the range *BC*, showing that the film thickness has reached a minimum value and approach of the surfaces has ceased. If rotation is stopped at time t_1 and resumed at stated intervals only long enough to take a torque observation, it will be found that the torque is progressively increasing. After rotation has been stopped for some time, it will also be found that a slight angular movement of the upper test surface will move the torque indicator from zero and it will not return, thus showing rigidity in the film. That the rigidity increases is shown by the continuous increase in torque with time over the range *CD*. If at time t_2 continuous rotation is resumed, the torque will drop rapidly at first and then more slowly until the previous constant value is reached at *E*. Rotation may be continued indefinitely over the range *EF* with no variation in torque, unless there is a change in film viscosity, due to variation of room temperature.

Fig. 9 shows the results of a continuous run of 52 hr with olive oil at 11.06 rpm and 40 lb per sq in. load, followed by periods of building up and removal of rigidity. Observations of torque versus speed at any time during the constant-torque period fall

on a straight line passing through the origin, provided the speed range is low enough to avoid frictional heating of the film. Fig. 10 shows such a curve taken at 50 hr in the test shown by Fig. 9. When rigidity has not been entirely removed by rotation, the torque-speed curve may be a straight line or a curve which, at zero speed, gives a value of torque proportional to the rigidity. It is obvious that torque observations with rigidity present will not be consistent since shear reduces rigidity, thus reducing the torque. There is, however, some minimum speed below which rigidity will not be entirely removed. No attempts were made to discover this limiting speed, but it was observed with olive oil at 40 lb per sq in. load that rigidity would build up very slowly indeed at 0.9 rpm.

After these tests the surfaces appeared to be exactly the same as at the start. There were no traces of scratching or metallic contact even at high magnifications. The first indication that the film will carry small electrical currents occurs at calculated film thicknesses of about $(10)^{-4}$ in. Before constant-torque conditions are reached, the circuit with the film in series carries the same current as when the film is short circuited by a copper wire. Rigidity does not appear to influence the resistance of the film.
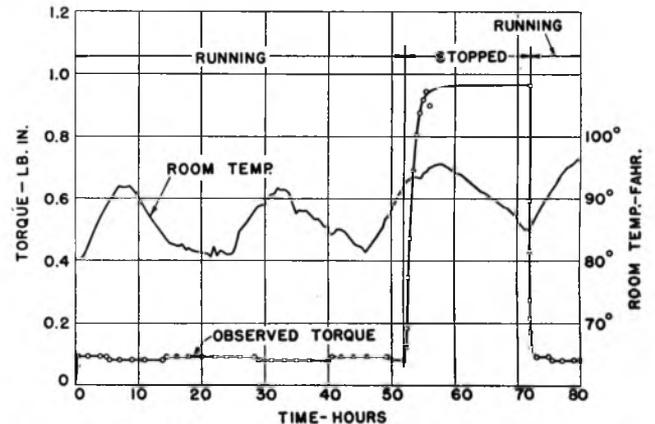
FIG. 9   EIGHTY-HOUR TEST WITH OLIVE OIL; TORSIONAL DISK VISCOMETER
(Plane disks, $^7/_8$ in. diam; 11.06 rmp; load, 40 lb per sq in.; average temperature, 87.9 F; $\mu = 7.80(10)^{-6}$ lb sec per sq in.)
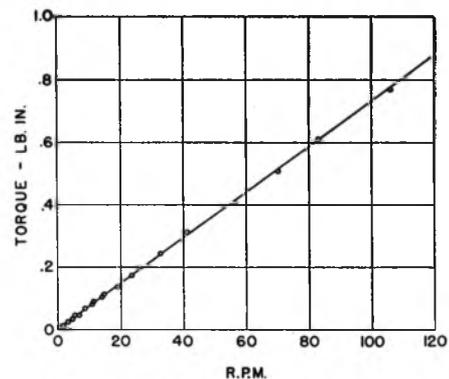
FIG. 10   VARIATION OF TORQUE WITH SPEED; TORSIONAL DISK VISCOMETER
(Plane disks, $^7/_8$ in. diam; load, 40 lb per sq in.; data taken at 50 hr in test shown in Fig. 9.)

Previous reference has been made to the excess oil from the very thick starting film, flowing out under the applied load and forming a seal around the edge of the film. The meniscus of this seal is in such direction as to cause a slight tension at the edge of the film. This is of little importance except to point out that

there is no pressure restricting flow; in fact, the results are the same if all excess oil is removed.

To measure the final film thickness a $3/32$-in-diam pin with a spherical babbitt tip was fitted into a hole in the lower disk, as shown in Fig. 11. The descending upper plane forces the pin into the hole and the height of its tip above the lower plane is measured after the test by counting the interference fringes (13) between the surface and a small optical flat in intimate contact with an edge of the plane surface and the pin. Film thicknesses thus measured after tests with several oils at loads up to 800 lb per sq in. are shown in Fig. 12. For example, with Atlantic superheat cylinder oil and a load of 300 lb per sq in., the torque reached
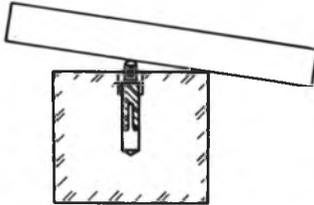
FIG. 11 PIN METHOD OF MEASURING MINIMUM FILM THICKNESS
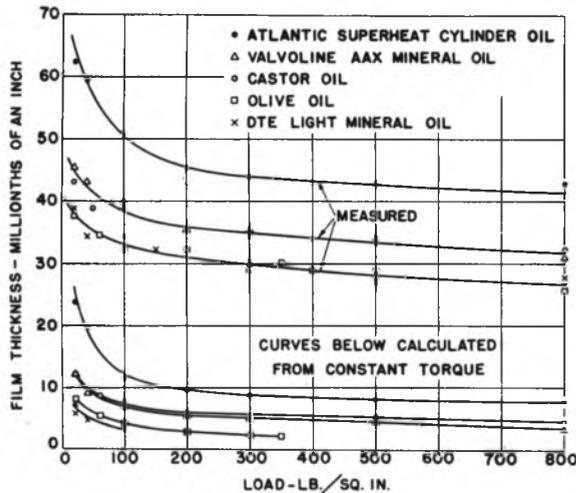
FIG. 12 MEASURED AND CALCULATED MINIMUM FILM THICKNESS WITH VARIOUS OILS; TORSIONAL-DISK VISCOMETER

(Plane disks, $7/8$ in. diam; 11.06 rpm. Atlantic superheat cylinder oil; average temperature, 72.4 F; $\mu = 670(10)^{-6}$ lb sec per sq in. Valvoline AAX mineral oil: average temperature, 71.2 F; $\mu = 113(10)^{-6}$ lb sec per sq in. Castor oil: average temperature, 72.2 F; $\mu = 115(10)^{-6}$ lb sec per sq in. Olive oil: average temperature, 73.6 F; $\mu = 11.04(10)^{-6}$ lb sec per sq in. DTE light mineral oil: average temperature, 72.4 F; $\mu = 9.42(10)^{-6}$ lb sec per sq in.)

a steady condition one hour after the load was applied and remained constant during the next hour. The test was then discontinued, the surfaces separated, washed with carbon tetrachloride and finally with alcohol. The optical flat showed the central pin to be $44(10)^{-6}$ in. high and this is the final constant film thickness under the test conditions. From the known bulk viscosity of the oil and the observed torque, the calculated film thickness from Equation [11] is found to be $8.8(10)^{-6}$ in. Variation of speed shows the torque to be directly proportional to the speed and there is no rigidity in the film in the direction of motion as long as rotation is continued. The lubricant is therefore behaving as a purely viscous fluid in so far as the direction of rotation is concerned. From the observed torque and the measured film thickness, the effective viscosity in the direction of motion has been increased 44/8.8 or five times the ordinary bulk viscosity. In the radial direction, however, the effective viscosity has been apparently increased to infinity, since no oil will flow from the film

$44(10)^{-6}$ in. in thickness supporting a load of 300 lb per sq in. This may be explained by the fact that rigidity will build up unless the rate of shear is above some critical value. The speed of rotation is sufficient to prevent the formation of rigidity in the direction of motion except perhaps to a small distance from the center. The rate of shear in the radial direction, however, is due to radial oil flow only and this decreases as the surfaces approach. Apparently when the rate of shear in the radial direction falls below the critical value, rigidity rapidly builds up, thus stopping further radial flow from the film. The surfaces are therefore unable to approach and a minimum film thickness is established.

Preparations for the foregoing tests were made under ordinary room conditions; hence, we may not exclude the possibility of contamination from the atmosphere. It is also probable that the commercial lubricants used were not chemically clean. A mass of particles of undefined nature would hold the surfaces apart, but the experimental evidence is against this possibility. If the particles were hard enough to hold the surfaces a definite distance apart the intensity of pressure on the larger particles undoubtedly would be great enough to cause scratching. If the particles were soft enough to change shape under load, they would also be expected to break down under continuous motion and permit closer approach of the surfaces. This would be evidenced by increasing torque. The presence of particles would also tend to increase friction to values well above those observed. Probably the most convincing argument against surface separation by foreign particles is the fact that, immediately after rotation has ceased and before rigidity has become appreciable, if the torque indicator is moved from its zero position, it will immediately spring back. This would be impossible if the surfaces were separated by particles under sufficient pressure to support the load. It seems beyond question that the surfaces are not held apart by particles in the film and any particles present must be smaller than the distance between the surfaces.

The fact that time is required to build up rigidity in the film and time is also required to remove it by motion suggests molecular rearrangements in the film. In such cases, it is reasonable to expect that both building-up and tearing-down actions would be rapid at first and then proceed more slowly as constant conditions are approached. Torque observations indicate that some such phenomenon actually occurs. It would also be expected that the building up and removal of rigidity would be more rapid in the less viscous fluids where molecular motion is more free. This is also observed to be the case. But the less viscous the oil the thinner the film under a given load, and presumably surface influence would become more pronounced as the film thickness is reduced.

Referring again to Fig. 12, the calculated values of film thickness were based on the known bulk viscosity of the oil and the constant torque along the range BC or EF of the curve in Fig. 8. At loads up to 350 lb per sq in., constant-torque conditions with olive oil were reached in which the friction coefficients were of the order of 0.007. At loads above 350 lb per sq in. with olive oil, and above 100 lb per sq in. with DTE light, no constant-torque conditions could be found. For the first few minutes of the test, the torque followed the theoretical law and then gradually increased to comparatively high values where it remained unsteady, fluctuating between values giving friction coefficients from 0.15 to 0.30. Measured film thicknesses under these conditions, however, fell on a smooth curve consistent with previous measurements at lower loads and constant-torque conditions. It is interesting to note that lard oil, which has about the same viscosity as olive oil, carried a load of 800 lb per sq in. under constant-torque conditions. Oleic acid and a mineral oil of somewhat higher viscosity than DTE light also carried the 800-lb per sq in.

load under constant-torque conditions. This is also true of white Russian mineral oil, the latter presumably nonpolar. Addition of 2 per cent of an ester of phosphoric acid to the above mineral oil made no difference whatever in torque or final film thickness.

Upon separation of the surfaces after a test under constant-torque conditions, the residual films appeared to be evenly spread over the entire area and of variable color depending somewhat upon the angle of view. Immediately after separation, castor- and olive-oil films showed the effects of surface tension by drawing up into tiny circular droplets. These arranged themselves in a geometrical pattern not unlike honeycomb but lacking its regularity of shape. About 10 min were required for the complete formation of the pattern. The final appearance was that of a piece of spider web dropped over the surface. Russian
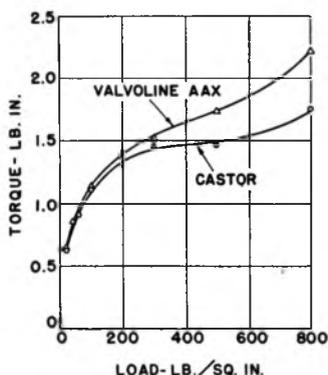
FIG. 13   VARIATION WITH LOAD OF CONSTANT TORQUE AT A GIVEN LOAD; TORSIONAL DISK VISCOMETER
(Plane disks, $^7/_8$ in. diam; 11.06 rpm; $\mu = 115(10)^{-6}$ lb sec per sq in.)

mineral oil and lard oil displayed similar tendencies. The pattern formed by oleic-acid film was much finer than the others and much more rapidly arranged. The commercial mineral lubricating oils showed no similar tendencies; the residual films remaining fixed in the presence of air and thoroughly wetting the surfaces.

In tests with olive oil at the heavier loads, where no constant-torque conditions could be found, the forces set up by rotation were strong enough to tear the films away from the surfaces and break them up into tiny white fragments. These fragments were so small that in some cases the separated surfaces had the appearance of being streaked with paraffin wax. At high magnification, the tiny white fragments were seen to be spots of oil exhibiting characteristic color fringes at their edges, but instead of being spherical segments like the droplets formed by surface tension, they were of irregular shapes with three to eight sides. Generally some round droplets were also present. Even though the film had broken up, the surfaces invariably showed no signs of scratching or wear. Apparently metallic contact had not occurred even though the friction coefficients were as high as would be expected with unlubricated surfaces. Similar tests with DTE light mineral oil generally resulted in faintly scratched surfaces or circular bands visible because of a higher polish than the remainder of the surfaces. Apparently the mineral oil was not as clean as the olive oil and the particles present were hard enough to scratch in the former case and soft enough to give a polishing effect in the latter. The films showed no tendency to break away from the metal as in the case with the olive oil.

Since the ordinary bulk viscosities of castor oil and Valvoline AAX are closely the same at the test temperature, and the measured film thicknesses are about the same, it would be expected that the same torque would be found at the same loads and speeds with the two oils. From Fig. 13 this is seen to be true at low

loads but, as the load is increased, the mineral oil gives greater friction than castor oil. At 800 lb per sq in. the friction with the mineral oil is approximately 27 per cent above that with castor oil. Increased viscosity due to film pressure (12) would account for about 8 per cent greater friction with the mineral oil. Corrections have been made for the slight differences in room temperature at which the various tests were made.

DISCUSSION AND CONCLUSION

Results of part I indicate that under very light loads a minimum film thickness is reached which is about 30 to 40 millionths of an inch. The order of these measurements is substantiated by the direct measurement at heavier loads described in part II. It now appears that the apparent increase in film thickness observed overnight in part I might possibly be due to the building up of film rigidity. From the results of part I, it appears that, at film thicknesses of about $(10)^{-4}$ in., the electrical resistivity of the film began to decrease. At thicknesses of about $60(10)^{-6}$ in. the film carried small currents as readily as the copper wire in the circuit. This also agrees with similar observations in part II.

The comparatively simple technique of the experiments described in part II and the ease with which results may be reproduced, offer an attractive method for further study of friction in thin films. The method is free from the complications and uncertainties introduced by abrasion and wear. An improved method of measuring film thickness is desirable.

It has been shown that, when two optically smooth plane steel surfaces are brought together in the presence of oil in such manner that the formation of a parallel film is possible, the distance between the surfaces will reach some definite minimum value depending upon the viscosity of the oil and the external force acting upon the surfaces. At room temperature, with surfaces $^7/_8$ in. diam, this minimum film thickness varies from about $30(10)^{-6}$ in. with olive oil to about $60(10)^{-6}$ in. with a heavy cylinder oil. The minimum film thickness is not greatly changed by loads up to 800 lb per sq in. of surface area.

The nature of this separating film is not known but experimental evidence indicates that the proximity of the metal surfaces has a profound influence upon the forces normally acting between the molecules of the liquid. This gives the film a high degree of mechanical stability, as evidenced by its ability to support considerable load without radial flow. The film also assumes the curious property of directional rigidity which seems to depend upon the direction, rate, and duration of shear.

At the present time there is no direct experimental evidence that rigidity is due to oriented molecules. But the rate at which rigidity builds up and the rate at which it is removed by shear points to the possibility of molecular rearrangement. If molecular orientation is due to surface influence it would be expected that the plane of least influence would be midway between the surfaces. The molecules at the center would be the first dislodged by shear thus causing a rapid reduction in friction. Resistance to dislodgment would increase with each molecular layer removed since the orientating forces would become stronger as the surfaces are approached. This would be evidenced by a decrease in the rate of torque reduction. Eventually a level would be reached where the orientating forces holding the molecules in place would be sufficiently strong to balance the shearing forces and dislodgment of molecules would cease; hence, the observed condition of constant torque at calculated film thicknesses well below the actual distances between the surfaces.

From the foregoing may be visualized a film composed of a fixed layer or layers of molecules firmly attached to each surface, and a central layer in which the molecules are kept free of surface influence by continuous shear. The central layer is there-

fore in the same molecular state as the liquid in bulk as far as the circumferential direction is concerned. Direct experimental proof of the existence of rigidity in the radial direction is lacking. But the fact is that, if the liquid in the central part of the film is free, as indicated by absence of circumferential rigidity, it does not flow radially under the external load. Hence, there must be some restraining radial force.

It was found that at loads above 350 lb per sq in. with olive oil and above 100 lb per sq in. with DTE light, no constant low-torque conditions could be established at 11.06 rpm. It would be interesting to investigate the effect of speed on these limiting loads. In fact, an explanation of the breaking up of films under loads too great to permit constant low-torque conditions would be most interesting indeed.

Liquids other than lubricants should be investigated for film-forming characteristics. Olive oil and DTE light should be highly filtered and the experiments repeated. Filtering should be done by the methods employed by Bulkley and also by centrifugal methods, since there is a possibility that Bulkley's methods might remove the components responsible for rigidity. It would also be interesting to discover if the diameter of the plane surfaces has any influence on final film thickness.

Finally, a study of polished metal surfaces would undoubtedly result in improved methods of producing smoother and truer finishes.

ACKNOWLEDGMENTS

BIBLIOGRAPHY

1 "A New Oil-Testing Machine and Some of Its Results," by A. Kingsbury, Trans. A.S.M.E., vol. 24, 1903, p. 144.

2 "Studies in Adhesion—I," by W. Hardy and M. Nottage, Proceedings of the Royal Society of London, series A, vol. 112, 1926, p. 64.

3 "Studies in Adhesion—II," by W. Hardy and M. Nottage, Proceedings of the Royal Society of London, series A, vol. 118, 1928, pp. 225–226.

4 "The Analysis of Commercial Lubricating Oils by Physical Methods," Lubrication Research Technical Paper No. 1, Department of Scientific and Industrial Research, London, 1930, p. 2.

5 "The Mechanism of Lubrication—II; Methods of Measuring the Property of Oiliness," by R. E. Wilson and D. P. Barnard, 4th, Journal of Industrial and Engineering Chemistry, vol. 14, 1922, p. 683.

6 "Viscous Flow and Surface Films," by Ronald Bulkley, U. S. Bureau of Standards Journal of Research, Research Paper No. 264, vol. 6, 1931, pp. 89–112.

7 "Physical Properties of Surfaces. II—Viscous Flow of Liquid Films. The Range of Action of Surface Forces," by S. H. Bastow and F. P. Bowden, Proceedings of the Royal Society of London, series A, vol. 151, no. A872, 1935, pp. 220–233.

8 "On the Range of Surface Forces," by F. P. Bowden, Physikalische Zeitschrift der Sowjetunion, vol. 4, 1933, p. 189.

9 "On the Theory of Lubrication and Its Application to Mr. Beauchamp Tower's Experiments, Including an Experimental Determination of the Viscosity of Olive Oil," by O. Reynolds, Philosophical Transactions of the Royal Society of London, vol. 177, part 1, 1886, pp. 157–234. "Papers on Mechanical and Physical Subjects," by O. Reynolds, University Press, Cambridge, vol. 2, 1901, p. 228.

10 "Versuche über die scheinbare Adhäsion" (Researches on Apparent Adhesion), by M. J. Stefan, Sitzungsberichte der Mathematisch-Naturwissenschaftlichen Classe der Kaiserlichen Akademie der Wissenschaften, vol. 69, part 2, nos. 1 to 5, 1874, pp. 713–735.

11 "Heat Effects in Lubricating Films," by A. Kingsbury, Mechanical Engineering, vol. 55, 1933, pp. 685–688.

12 "Influence of Pressure on Film Viscosity in Heavily Loaded Bearings," by S. J. Needs, Trans. A.S.M.E., vol. 60, 1938, paper RP-60-7, p. 348.

13 "Interference Methods of Standardizing and Testing Precision Gage Blocks," by C. G. Peters, and H. S. Boyd, Scientific Papers of the U. S. Bureau of Standards, Paper No. 436, vol. 17, 1922, pp. 681.

Discussion

C. H. BIERBAUM.[4] After reading the paper, the thought seems to persist that the oils tested were oxidized, at least so in a limited degree and that, therefore, the researches of Langmuir[5] seem to have a direct bearing. His findings are that the oil molecules are columnar in form and that, in the oxidizing process, one end remains hydrophobic and the other end to which the oxygen is attached becomes hydrophilic or metallophilic, so that these molecules will orient themselves perpendicularly upon a metal surface.

It was the writer's observation that by oxidizing an ordinary motor oil by bubbling pure dry air through it at the temperature of boiling water, for say 2 hr, the ultimate load-carrying capacity of this oil was increased by more than 4000 lb per sq in. of projected bearing area. This effect seems entirely due to the increased tenacity with which the oriented oil molecules attached themselves to the metal surfaces, an adsorbed effect.

Oils oxidized in the foregoing manner have a distinctly corrosive[6] action, in varying degrees, upon the different metals ordinarily used for bearing purposes. It seems safe to assume that the chemical affinity of such modified oils is greater for the more corrosive materials than for the less corrosive; as an illustration, silver, a more noble metal, is not corroded in this manner, which is one reason why in the pure state it has not given a better account of itself for bearing purposes.

Fortunately, it has been found that the full benefit of this oxidizing of oils can be taken advantage of without having the deleterious corrosive effect and, at the same time, be made into a positive inhibitor of corrosion, by the elimination of the corrosive by-product.

It would be interesting to know whether any oxidized oil molecules were present in the tests described in the paper; especially so with the mineral oils, since "the residual film remained fixed in the presence of air and thoroughly wetted the surfaces." Very interesting results can be expected in future tests with different materials for the two bearing surfaces.

The electrical conductivity of the film seems in keeping with the experience in the breaking down of polarized oil in electric transformers, a strong suggestion that oxidized molecules were present.

The fact that the addition of 2 per cent of ester of phosphoric acid had no effect either on the torque or film thickness in the experiments described seems in itself unaccountable, except for the possible presence of an oxidized oil film. The results seem to show that boundary film lubrication may be divorced completely from the properties of the lubricant as a whole.

High-power photomicrographs of the bearing surfaces before and after the tests would have been desirable.

Mr. Needs is certainly to be congratulated upon the accuracy and refinement of his tests. His paper shows distinct progress on the subject of lubrication.

L. J. BRADFORD.[7] The explanation advanced by the author for the difference between the calculated and measured film

---

4 Vice-President and Consulting Engineer, Lumen Bearing Company, Buffalo, N. Y. Life Mem. A.S.M.E.

5 "Constitution and Fundamental Properties of Solids and Liquids," by Irving Langmuir, Journal, American Chemical Society, vol. 39, 1917, p. 1848.

6 "Corrosion Effects of Lubricants Upon Bearing Surfaces," by Christopher H. Bierbaum, The Iron Age, August 31, 1933.

7 Professor Machine Design, Pennsylvania State College, State College, Pa. Mem. A.S.M.E.

thicknesses shown in Fig. 12 of the paper is attractive, but not wholly convincing. We should expect the extent to which the molecules were detached from their static positions to be dependent upon the force with which they were held, and upon the distorting force applied to them. This latter will in turn be proportional to the rate of shear at the boundaries of the active portion of the film. If we assume that the viscosity of the oil in the active portion is the same as that of the oil in bulk, we may then take the distorting force as inversely proportional to the ordinates of the curves of calculated film thickness. It would be greatest at heavy loads and least at light loads. We should, therefore, expect the static, oriented portions to be thickest at the light loads and thinnest at the heavy.

An examination of the curves shows that this is true. The difference between the measured and calculated thicknesses, which we may take as representing twice the thickness of the oriented, static layer, is greatest at the light loads and least at the heavy. We should, however, expect the orientation to be more vigorous with a substance which is decidedly unsymmetrical in molecular structure, than one with greater symmetry.

The observations recorded in Fig. 12, do not sustain this expectation, for curves for Valvoline, a mineral oil and castor oil are almost identical, while that for Atlantic superheat cylinder oil shows a markedly greater difference between the measured and calculated thicknesses. Apparently viscosity is still associated in some way with these phenomena.

Figs. 8 and 9, showing marked increases in the torque if the apparatus is allowed to stand, suggest that the orientation takes time to form, but that it breaks down quite rapidly.

F. J. Villforth and the writer repeated the stop-and-start experiments, using a tapered-plug viscometer and olive oil. Clearances much greater than those mentioned in the paper were employed, in order to get some idea of the distance to which these surface effects were felt.

With a clearance of 0.0055 in. an increase in torque of about 10 per cent was found when the apparatus was allowed to stand for several hours. When the clearance was reduced to 0.0033 in. the increase was found to be about 20 per cent. A further reduction to 0.00164 in. raised the increase to about 25 per cent. These results were obtained with rates of shear of 786, 978, and 1950 reciprocal sec, respectively. There was also some indication that after a rest the torque first increased and then decreased to a value below that at the start.

If the foregoing results can be validated by subsequent and more extensive experimental work, they would indicate that surface effects extend to distances which are truly enormous in terms of molecular dimensions.

RONALD BULKLEY.[8] The methods of test described in this paper are refreshingly novel and possess great possibilities for the further elucidation of thin-film phenomena. In attacking related problems in the past, several investigators have drawn conclusions regarding the nature of liquid films as close as a few millionths of an inch to a solid wall, but in most cases their actual experiments have been made on films whose thickness has been perhaps 200 times this amount. Now, the author has been able to work with actual films only 0.00002 or 0.00003 in. in thickness and truly this is a great stride forward in the art of thin-film experimentation. His conclusions, however, are not in accord either with classical theory or with the conclusions of several previous experimenters in the field.

The concept of long-range forces which increase the viscosity of liquids out to distances of 0.0001 in. or more was championed 15 or 20 years ago by Sir Wm. B. Hardy. In 1922, Wilson and

Barnard published experiments purporting to show that oily liquids of polar characteristics, e.g., oleic acid, could be solidified by these forces clear out to distances of 0.005 or 0.006 in. from a solid wall. For 8 or 10 years these results were accepted and widely quoted by many lubrication engineers. Orthodox physicists and chemists pointed out, however, that on the basis of their theoretical calculations, molecular influence can extend outward to not more than about 0.0000005 in.

In 1931, the writer repeated Wilson and Barnard's experiments and discovered that the clogging of capillaries which they attributed to solidification of the liquid was due to lint and foreign matter in the oil. In the same year two of Hardy's own students, Bastow and Bowden, published a paper,[9] giving the same explanation for Hardy's anomalous long-range results, namely, dust and lint in the liquid. In a paper 2 or 3 years later, the same men described viscosity experiments conducted in an apparatus not too different from the author's disk viscometer. These experiments completely confirmed the writer's earlier work, indicating, as he had claimed, that no rigid layer existed on a solid wall out to distances greater than 0.000001 or 0.000002 in.

It is difficult indeed to reconcile these results with those now found by the author. Another difficulty arises from the fact that even Hardy and Wilson and Barnard did not postulate or find any orientating influence for ordinary hydrocarbons or pure petroleum oils, but only for polar materials or materials possessing good oiliness, such as lard oil or oleic acid. The author, however, finds just as pronounced an effect for a mineral oil as for castor oil and olive oil.

Moreover, his conclusions do not seem always to be justified by the data he presents. For example, of the 12 runs he made on 3 oils in the disk viscometer, 5 agreed closely with the theoretical curves. The other 7 were wide of the theoretical curves and the author notes as a general comment that "satisfactory duplication of tests was found impossible." Yet notwithstanding the narrow margin between the number of curves which did and those which did not agree with theory he concludes, contrary to theory, that "the above experiments indicate that there is some limiting film thickness reached by plane surfaces approaching under load." He places the thickness of the limiting film reached under these conditions at about 0.00003 or 0.00004 in.

In the second viscometer, much higher loads are used and the limiting film thickness is, therefore, reached in a much shorter time. The thickness of the film, however, is substantially the same as in the first viscometer, notwithstanding the fact that the loads may be some 4000 times as great. It is hard to imagine what kind of material the oil has been converted into in this relatively thick layer to permit the layer to remain as thick under 800 lb per sq in. attempting to squeeze the oil out as under only 0.2 lb per sq in. The picture the author draws calls for the intensification of viscosity to be greatest at the wall where the forces are greatest and to diminish as the distance from the wall is increased until at last out in mid-stream, 0.000015 or 0.00002 in. away, the viscosity is the normal bulk viscosity. Would we not expect heavy loads to squeeze out more of these partially solidified layers than light loads and thus to give a thinner film? Maybe we could not expect the light load to give a film 4000 times as thick as the load which is 4000 times as great, but should we not expect the film under the light load to be, say, 1000 times as thick or maybe 100 times as thick or at least 10 times as thick? In the author's experiments, the films are of the same thickness regardless of load. This seems to suggest some limitation, not yet discovered, inherent in his apparatus or method

[8] Haddon Heights, N. J.

[9] "On the Contact of Smooth Surfaces," by S. H. Bastow and F. P. Bowden, *Proceedings* of the Royal Society, vol. 134. 1931. pp. 404–413.

of experimentation which leads him to erroneous conclusions.

Failing, to date, to find any such limiting conditions, the author's only other recourse, naturally, has been to invent a new concept to explain his data. This has taken the form of a layer of material some 500 or 1000 molecular lengths in thickness which possesses the surprising property of being a solid in one direction and a simple viscous liquid at the same time in another direction.

Regarding this revolutionary concept of directional rigidity, it should be observed that every investigator has a right to invent new concepts when his data are incapable of explanation in terms of old concepts. The author has found it necessary to invent the concept of directional rigidity in liquids which up till now have been regarded as possessing uniform properties in all directions. Until some one can explain his results in a more simple or more conventional manner his new concept must stand.

J. R. Connelly.[10] The first part of this discussion will deal with specific matters mentioned in the paper and the second will be a brief statement concerning some of the larger questions in the field in which this investigation exists. In the course of the writer's comments some questions will be raised that cannot be answered and some techniques of the author questioned where an improvement cannot be suggested at present. The specific discussion follows:

1 The method of indicating contact between the micrometer and the beam is open to question because of the possible unevaluated constant and variable errors. The writer and his associates have struggled with this problem for some years and do not feel they have a satisfactory solution, although this method has been tried.

2 Which results were obtained using normalized low-carbon steel and which using hardened tool steel? It is often contended that a low-carbon steel, work-polished against a soft bearing, will react differently from a hardened tool steel which is not work-polished. It seems that some future investigation by this method should endeavor to compare a work-polished (probably burnished is a better term) low-carbon steel with a hardened tool steel. It might also be worth while to prepare two highly polished flat glass disks, since the surfaces exposed would not furnish recesses for the lubricant.

3 In the section devoted to disk-viscometer experiments the author mentions that "satisfactory duplication of results was found impossible." For the castor oil, the variability of results is explainable and, for the olive oil, the extreme variation of one determination is more of quantity than kind. However, in the case of the DTE light mineral oil, it may be said that agreeing data are definitely in the minority. It raises the question of whether all the variables are under even approximate control.

4 In Fig. 10, the author shows a plot of torque taken during the constant-torque period, against rpm. In discussing this plot, the statement is made that the relation is a straight line passing through the origin, provided the speed range is low enough to avoid frictional heating of the film. Now the writer does not wish to contend whether the relation is a straight line or not but to infer that a rotative speed of 100 rpm does not produce frictional heating of the oil film seems far-fetched indeed. Some other explanation would be more satisfactory.

In discussing the method used in the investigation, the author states that it "is free from the complications and uncertainties introduced by abrasion and wear." This statement is true as far as it goes, but to it could be correctly added, "and gives no information on abrasion and wear."

It would seem that our knowledge has advanced to a state where we can recognize a number of different classes of conditions in bearing operation, as follows:

Class A dry rubbing contact, resulting in reduction of contact surfaces (abrasion).

Class B rubbing contact in the presence of a lubricant, resulting in reduction of contact surfaces (wear).

Class C relatively thin film separating the surfaces (present paper).

Class D relatively thick film separating the surfaces (hydrodynamic theory).

The term boundary film or boundary lubrication would seem to apply to both classes B and C in the sense that they are intermediate between dry-rubbing and thick-film lubrication. For purposes of study it would seem wise, however, to keep these classes separate in our thinking since the variables give evidence of being quite different.

M. D. Hersey.[11] While a variety of new phenomena are reported in these experiments, they can be harmonized somewhat by noticing the close analogy which they bear to the more familiar phenomena of plastic deformation in everyday tests of greases, and many other types of plastic material, particularly materials that are strongly thixotropic. This analogy might be made more complete if the author could supply a typical curve for torque against speed, taken during the recovery period after rigidity has set in.

It would seem that the possibility has not been entirely ruled out that the observations may, in part, be explained as due to a slight degree of plasticity caused by suspended particles; not necessarily foreign particles. Such plasticity might be present in commercial oils without becoming noticeable until the film is sufficiently thin.

Two experiments might be suggested for the future: (1) Repetition of some of the rotation tests, substituting annular surfaces to simplify the mathematical analysis; and (2) observations of thixotropy of any convenient plastic material in bulk to see if the properties in one direction are influenced by motion at right angles.

G. B. Karelitz.[12] The work reported in this paper is so extraordinary that one is tempted to look for signs of experimental error. It seems, however, that none of the measurements or techniques can be objected to. Yet it is difficult to believe that the influence of steel on oil molecules would penetrate to a depth of 0.00002 to 0.00003 in. It would mean that the effect of the contact surface may be felt at a distance of 200 to 500 molecular dimensions from the steel surface. In a number of experimental bearings, such as tested by McKee at the Bureau of Standards, the minimum film thickness was of the order considered in the paper, 40×10^{-6} to 60×10^{-6} in. It appears that these bearings should have shown a high coefficient of friction, but they did not. It may be that the agitation in an actual bearing, together with higher speed, would prevent the stratification of the lubricant into an orientated layer.

If the orientation of the molecules extends that far into the film, the several layers near the metal surface must be lined up rather well. It should then be comparatively easy to detect this quasi-crystalline structure by X rays, if pyrex-glass disks are used instead of steel. Reasonably flat disks could be pressed together, with an oil film trapped between them and a pencil of X rays played through the assembly. Of course, it may be that

glass does not produce the same orientation as does steel. A number of observations with techniques different from that used by the author would be required before the existence of directional rigidity could be accepted. This is said not to minimize in any way the importance of the contribution but rather to emphasize the significance of the report.

The low electrical resistance of the film is also very disturbing. It is very unusual that a film 10^{-4} in. thick and only a fraction of a square inch in area, should pass a detectable electric current with an imposed voltage of 25 mv.

One can only hope that this investigation will be followed up by other experimenters who will have the opportunity to study the physics of a thin oil film. The painstaking manner in which the work was performed and reported by the author speaks for itself and adds further credit to the laboratory of Dr. A. Kingsbury.

MELVIN MOONEY.[13] Regardless of what the explanation may be for the formation of a stable thin film of oil, the facts established in the author's experiments are of fundamental importance in problems of lubrication. The explanation given for the facts observed is not very convincing. The data can readily be explained on the assumption of a layer of foreign particles between the parallel plates. Such particles would not be eliminated by centrifuging if they have approximately the same density as the oil. When the upper plate is rotated, the particles roll and form effectively a ball or roller bearing. If they are irregular in shape or slightly compressible, they would tend to increase the separation between the plates when the upper plate is rotated at a fixed pressure. This appears to be in agreement with the experimental facts. (The data on this point are not quite conclusive, for they require an extrapolation of the thickness-pressure curves under rotation back to 0.2 lb per sq in. pressure.) On the other hand, if the strength and rigidity of the final film depend upon a special arrangement of the molecules of the oil, it would be anticipated that the shearing action, when the plates are rotated, would disturb the arrangement and cause a decrease in the thickness of the film. This is contrary to the observed facts. The rigidity to rotation developed on standing may be ascribed to yielding and flattening of the foreign particles.

The difficulty of reconciling the author's explanation of his own experiments with the results obtained by Bulkley with fine capillary tubes is suggested but not discussed by the author.

It is to be hoped that the work with the parallel plates will be continued as planned. There is a variation in the experimental procedure which might be adopted to determine definitely whether the rigidity of the film is due to properties of the oil itself or to fine particles suspended in the oil. The variation here suggested is that one of the test plates be drilled through its center. Then, after a stable film has been produced, it would be possible to test whether oil under pressure can be forced through the central hole and out between the two plates. If so, the rigidity must be due to isolated points of support, presumably foreign particles. If not, the entire film is a rigid structure involving all the oil.

M. MUSKAT AND F. MORGAN.[14] Whether one agrees or disagrees with the observations presented in this paper, the importance of its implications in all problems of the flow of liquids near solid boundaries as well as in boundary lubrication must be recognized. In view of the perhaps almost universal opinion, based largely on the researches of Bulkley and of Bowden and his colleagues, that no such effects exist, the author and Dr. Kingsbury should certainly be commended for their activity in a field

in which the possibility of obtaining positive results must have appeared very small.

Before all the previous results are discarded, however, or at least explained on the basis of different experimental conditions, it is perhaps worth asking if the results of this paper can be interpreted in any other way. For instance, the possibility that dust particles may contribute to the experimental results has already been considered, but apparently no great pains have been taken to remove this unknown factor from the experiments. The elaborate precautions necessary to obtain clean surfaces and purified air or liquids relatively free from dust particles have been described by Bastow and Bowden.[9] In spite of the methods used, however, they report that, under the ultramicroscope, purified alcohol still showed occasional dust particles. In fact the presence of some sort of contamination in the author's experiments seems almost to be postulated from the observation in part I that the micrometer zero could only be set by bringing the surfaces together with platinum wires between them, since a false reading would probably be obtained due to surface contamination or adsorbed films if they were brought together without the wires.

In any case, even though the author does feel that dust particles were not present, it is interesting to observe that if we do assume their presence practically all of the experimental results can be explained without invoking the peculiar rigidity properties proposed in the paper. Thus the linearity of the friction-versus-speed curve would simply be the result to be expected of a normal viscous film. As the dust particles would be supporting the load, the surface-tension forces at the edge of the surfaces would more than suffice to keep the liquid in without requiring any rigidity characteristics. The increase in torque observed after the rotation has been stopped might be due to a flattening of the dust particles on standing and subsequent return to approximately spherical shape upon rolling between the rotating disks.

Admittedly, the discrepancy between the directly measured film thicknesses and equivalent film thicknesses calculated from the torque observations must then be ascribed to errors in the pin method of direct thickness measurement. But, until there are other independent confirmations of the accuracy of this method, it is difficult to have sufficient faith in it to compel acceptance of the rigidity hypothesis. As to the lack of wear and scratches, this might be explained if the dust particles act in a manner similar to that of a polishing agent, even though they apparently do not become broken up.

While this picture is presented only as a working hypothesis, the alternative interpretation of the author requires that the film have a tremendous anisotropy. Thus in the direction of rotation the rigidity is equivalent only to a viscosity 5 times its normal value, whereas its radial viscosity is practically infinite. Furthermore, when the disks are at rest, the film apparently has an isotropic rigidity. Moreover, if rotation tends to remove the rigidity, it is surprising that, over the entire range of speeds of 120 rpm practically to zero, the effective viscosity remains so strictly constant.

If these properties are substantiated, the author will certainly have established a most remarkable phenomenon. It is felt, however, that just because this phenomenon would have such a profound bearing upon the general problem of lubrication, it is highly important that all factors which may permit alternative explanations of the experiments be positively eliminated before what are generally believed to be well-established physical concepts are discarded.

B. L. NEWKIRK.[15] In this paper, the observed effects are so

[13] U. S Rubber Products, Inc., Passaic, N. J.
[14] Gulf Research & Development Company, Pittsburgh, Pa.
[15] Rensselaer Polytechnic Institute, Troy, N. Y. Mem. A.S.M.E.

striking that they call for careful study and further efforts to check them experimentally from new angles if possible. Particularly impressive is the consistency of measurements of the individual curves of Figs. 3, 4, and 5. The departures of individual points from smooth curves are only a few millionths of an inch. In Fig. 3 agreement with theory is close. Figs. 4 and 5 show discrepancies which would be reduced materially if some reason could be found for a change in the zero reading from one test to another. These curves show some apparently real diversities in shape that could not be so explained. For example, the curve marked by crosses in Fig. 5 would not fit the theoretical curve even after displacement. In view of the author's remarks about the difficulty of the zero measurement, it is not surprising that vertical displacements of the curves appear. A wave length of sodium (yellow) light is nearly 0.000024 in. On the other hand, it would certainly not be wise to dismiss these discrepancies in the view that they may be due to zero errors, for it is possible that some important fact would be brought to light if we could find the real cause. If means could be found to check the zero measurements independently, something further of importance might be discovered.

The experiments of part II also turn to a considerable extent on the measurement of distance between the plates. The method used in this case to measure the distance seems admirable but an independent check of a factor of such critical importance would be desirable if it could be managed. The measured distances between the plates are of the order of 400 times the length of the long polar molecules. The apparent fluidity for circular motion and rigidity for radial motion and the rigidity developing during periods of no rotation are very remarkable. If the surfaces had bumps and hollows or if they were distorted by pressure or the mounting, this might account for some of the observations, but the author's statement that the surfaces were flat to better than 0.000001 in. and the care taken in the design and mounting of the disks seem to exclude these possibilities.

K. TERZAGHI.[16] The personal experience of the writer concerning the subject of the paper is limited to the physical properties of boundary films in clay-water mixtures. Therefore the writer was much impressed by the similarity between his own experience with water films and that of the author with oil films.

The findings of the writer can be condensed into the following statements: If the average width of the voids in a clay-water mixture is smaller than about 0.1μ the average effective viscosity of the water increases rapidly and, at the same time, the water assumes more and more the characteristics of a solid substance.[17] The increase in viscosity is apparently due to the restraining influence of the proximity of the solid surface on the movement of the molecules of the liquid.[18] The existence of an almost perfectly constant ratio between the normal pressure and the corresponding shearing resistance of the boundary film seems to be due to an orientation of the molecules of the liquid within the boundary film.

In 1933, B. Derjaguin published the results[19] of investigations concerning the elastic properties of water films between an optically plane and a curved surface of glass. The thickness of the film was determined by observing the Newtonian rings around

the point of minimum distance between the two surfaces. The data required for computing the modulus of shear of the boundary film were obtained by recording the period of oscillations produced by the application and subsequent release of a torque. The following table contains the results of the investigation, quoted from the paper:

Thickness of film in microns	Modulus of shear in absolute units
0.089	190×10^6
0.093	170×10^6
0.137	4×10^6
0.150	0.0

This table shows that the thickness at which the film began to exhibit a measurable modulus of shear (0.14μ) is of the same order of magnitude as the thickness at which the writer noticed a marked increase of viscosity (0.1μ). However, as soon as the water acquires elastic properties, one can no longer call it a liquid and the physical meaning of the term viscosity changes. The transition from the liquid to the solid state, which takes place if the thickness of the film is decreased, also requires a corresponding progressive departure from Stefan's equation for the relation between the time and the thickness of the film (Equation [6] in the paper). In order to bring this departure more clearly into prominence, it would be advisable to plot the value $h\sqrt{t}$ against $\sqrt{t}$, because according to Equation [6] the value $h\sqrt{t}$ should be a constant, which appears in such a graph as a horizontal line. The departure from a horizontal line is more easily detected than that from a hyperbolic curve.

As h approaches a final constant value h_o the empirical curve approaches an asymptote through the origin $\sqrt{t} = 0$ and the value h_o is determined by the slope of this asymptote. The writer would appreciate it if the author would present such a graph in his final discussion.

The rapid increase of the torque after stopping, as shown in Figs. 8 to 10 of the paper reminds the writer of the thixotropic phenomena in the clay-water mixtures. The increase of stiffness with time under unaltered external conditions, characteristic of thixotropic phenomena seems to be due to the gradual building up of a molecular structure within the boundary films. Such an internal change may also account for the ultimate increase in the thickness of the films in the tests illustrated by Figs. 3 to 6.

As soon as the rotation ceases, the resistance against sliding approaches the value determined by the coefficient of static friction for lubricated surfaces, which is normal pressure per unit of area times coefficient of friction.

The coefficient of static friction for olive oil (Figs. 8 and 9) seems to be of the order of 0.14. For castor oil W. B. Hardy[20] gives a value of 0.10. In this connection, it is interesting to note the extraordinary difference between the rate at which the sliding resistance approaches the static value under a pressure of 40 lb per sq in. (Fig. 9, very rapid) and under 100 lb per sq in. (Fig. 8, very slow). Noticing this difference one feels tempted to consider whether an increase of the pressure may not increase the resistance against re-establishing the molecular structure within that part of the film which is kept by rotation in a liquid state.

Regarding the rigidity of the film during rotation, the writer is inclined to believe that the rigidity existed throughout the films, except within a layer with a thickness equal to the diameter of a few molecules along a plane half way between the solid surfaces. This may also account for the failure of the oil to flow out in a radial direction, in spite of liquid behavior with respect to rotation. Finally it may eliminate the apparent contradiction be-

[16] Harvard University, Cambridge, Mass.

[17] "Versuche über die Viskosität des Wassers in sehr engen Durchgangsquerschnitten," by K. Terzaghi, *Zeit. für Angewandte Mathematik und Mechanik*, vol. 4, 1924, pp. 107–113.

[18] "The Mechanics of Adsorption and of the Swelling of Gels," by K. Terzaghi, Fourth Colloid Symposium Monograph, Chemical Catalogue Company, Inc., New York, N. Y., 1926, pp. 508–708.

[19] "Die Formelastizität der dünnen Wasserschichten," by B. Derjaguin, *Zeit. für Physik*, vol. 84, 1933, pp. 657–670.

[20] "Note on Static Friction," by W. B. Hardy and J. K. Hardy, *The Philosophical Magazine*, London, vol. 38, 1919, pp. 32–48.

tween the measured and the calculated values of film thickness (Fig. 12 of the paper), because the calculation seems to be based on assuming that during rotation the film is liquid throughout its thickness. The writer would appreciate an expression of opinion by the author as to whether he considers this hypothesis compatible with his observations. In every other respect his conceptions regarding the structure of the oil films are practically identical with those of the writer, derived from a very different realm of empirical evidence.

AUTHOR'S CLOSURE

Many interesting points have been brought out in the discussion. Some of these points may be clarified, but in most cases the information is not available at present to answer the questions raised.

Attempts by Professor Bradford and Mr. Villforth to detect indications of film rigidity by means of a tapered-plug viscometer are most interesting. Since the rate of approach of the surfaces appears to be in agreement with theory, until the film thickness is close to its minimum value, it would be indeed surprising if surface influence were detected in such comparatively thick films as 0.005 to 0.006 in.

Dr. Bulkley passed highly filtered fluids through fine capillaries and arrived at the conclusion that there was no surface influence sufficient to produce a rigid layer out to distances greater than 0.000001 or 0.000002 in. from walls of platinum or glass. The author worked with films of oil of the order of 0.00003 in., bounded by two plane steel surfaces, and found unmistakable evidences of unexpected behavior. This behavior he attributed to surface influence. Aside from questions of preparation and purity of the liquids, technique, and others, which become apparent when reading the papers and comparing the findings, the two following important facts stand out: A single surface, such as the wall of a capillary tube, has little or no influence on a contiguous liquid. Two steel surfaces in close proximity exert an appreciable effect on the oil between them. These two findings are not essentially in conflict. All the experiments on which the present paper is based indicated that, when the films were relatively thick, their behavior could be calculated from simple theory. It was only when the two steel surfaces came close to each other that their influences became apparent. It was also found that there appeared to be an appreciable change in electrical resistivity as the films became thinner. This point was not investigated to the extent warranted by its importance, but it was noted that the beginning of the change in resistivity occurred at approximately the same time as the departure of the rate of approach from Stefan's law. Investigation along these lines might be fruitful.

Professor Connelly has pointed out that the work done in shearing the oil film must cause heating of the film. At low speed and low torque, however, the rate at which heat is generated is so small that its influence on the film cannot be detected with the present apparatus. The rate at which heat is generated depends upon work done in shearing the film, and not upon rotative speed alone. For example, in the test shown by Fig. 10 of the paper, the rate at which heat is generated at 100 rpm is about the same as at 30 rpm in the test shown by curve 2 in Fig. 14 of this discussion. Data are available to show that, under these respective conditions, heat is generated at a rate sufficient to cause a 10 per cent drop in torque in 1 min. Only a few seconds are required for a torque observation and, between observations, the speed is reduced to 11 rpm or less where heating of the film is not appreciable. During the constant-torque period in the test shown by Fig. 9 of the paper, the reduction in torque due to shearing of the film is probably not in excess of 2 per cent.

No difference in test results could be detected by using hard-
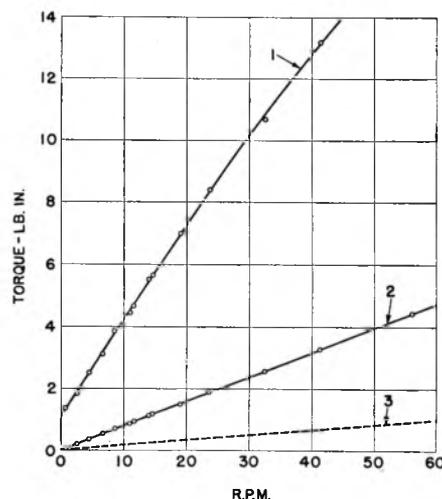
FIG. 14 VARIATION OF TORQUE WITH SPEED; TORSIONAL DISK VISCOMETER

(Plane disk, $^7/_8$ in. diam; load 100 lb per sq in.; castor oil, average temperature, 73.7 F; $\mu = 107(10)^{-6}$ lb sec per sq in.)
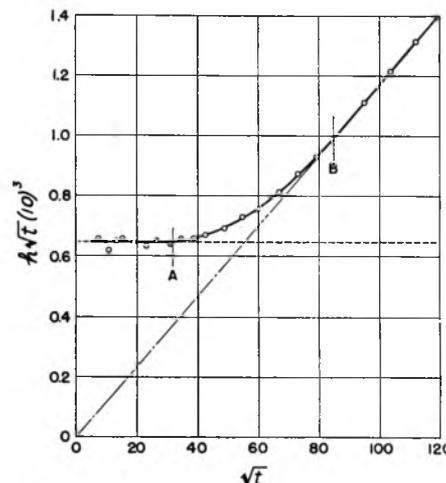
FIG. 15 TERZAGHI'S METHOD OF DETERMINING TIME OF DEPARTURE FROM STEFAN'S EQUATION; TORSIONAL DISK VISCOMETER

(Plane disk, $^1/_8$ in. diam; 11.06 rpm; load 40 lb per sq in.; castor oil, average temperature, 71.8 F; $\mu = 117(10)^{-6}$ lb sec per sq in.),

ened tool steel in place of normalized low-carbon steel. Preference of one steel over another for making the test surfaces was due only to the reasons stated.

The torque-versus-speed curves requested by Mr. Hersey are shown in Fig. 14 of this discussion. The data were taken with castor oil and a load of 100 lb per sq in. After a continuous run of 8 hr at 11.06 rpm, rotation was stopped and the film was allowed to remain undisturbed for $15^3/_4$ hr. Variations of torque with speed were then observed and plotted in curve 1. Rigidity at zero speed is shown by the intercept at about 1.05 lb-in. Continuous rotation was then resumed and when the constant torque of the previous day had been reached, the speed was again varied with the results shown by curve 2. As in all cases when rigidity is not present, the torque is proportional to the speed and the curve is a straight line passing through the origin. From Fig. 12 of the paper, the film thickness with castor oil at 100 lb per sq in. is $38(10)^{-6}$ in. and, for the ordinary bulk viscosity of the oil, the theoretical torque is shown by the broken-line curve 3. Comparison of the slopes of curves 2 and 3 shows that, in this film, surface influence has increased the effective viscosity to nearly 5 times its normal value.

That the data may be readily explained on the assumption of a layer of foreign particles between the parallel plates, as suggested by Drs. Mooney, Muskat, and Morgan, is not quite clear to the author. It is true that no elaborate precautions were taken definitely to exclude dust from the atmosphere during the short time required to apply the oil and start the test, but to visualize a particle that could account for the observed phenomena is very difficult indeed.

Dr. Terzaghi's suggestion that the point of departure from Stefan's Equation [6] may be more clearly seen by replotting the data, has been carried out in Fig. 15 of this discussion. Data therein presented are from a test with castor oil and a load of 40 lb per sq in., with continuous rotation at 11.06 rpm. The horizontal broken line represents the constant value of $h \sqrt{t}$, which in this case is 0.000648 from Equation [6]. The point of departure A is at $\sqrt{t} = 32$. Therefore, during the first 17 min, the surfaces are approaching at the rate calculated for a purely viscous fluid. Torque becomes constant at B, 2 hr after starting, and approach of the surfaces ceases. From the slope of the curve beyond B, the minimum film thickness is $11.7(10)^{-6}$ in.

This may also be found from the observed torque and Equation [11]. But from Fig. 12, however, we find the actual minimum film thickness is about $41(10)^{-6}$ in., nearly 4 times the theoretical value.

From Fig. 8 of the paper, it is seen that, with olive oil at 100 lb per sq in., the torque increases from 0.125 to 1.54 lb-in. over the 2-hr period CD. Time, in Fig. 9, is plotted in hours instead of minutes; hence, it is not quite so clear that, at 40 lb per sq in., the torque increases from 0.08 to 0.8 lb-in. in a similar 2-hr period. The rate of torque increase, therefore, is greater at the greater pressure and the final values reached depend upon pressure and time.

With reference to film rigidity and the general behavior of these thin films, there seems little difference between the conclusions which were reached by Dr. Terzaghi and those expressed by the author.

It is evident that further study of the boundary-film problem is desirable and warranted by the promise of additional information. This discussion has been most valuable in that it points out several paths for future research.

High-Speed Lightweight Trains

By C. T. RIPLEY,[1] CHICAGO, ILL.

The purpose of this paper is to outline the changes which have occurred during the last five years in high-speed passenger-train cars and in motive power for hauling them and the economic factors which have brought about these changes. The new designs for passenger cars and the materials used in their construction are discussed. A detailed comparison of steam-locomotive and Diesel electric-locomotive characteristics as they affect the operation of these new high-speed trains is presented. Test data are included to indicate the importance of comparative stress in track produced by two types of power. Reference is made to the steady improvement which has been made in steam-locomotive design, but it is shown that there is a need for some rather extensive experimentation to make this type of power more suitable for this particular class of service. In conclusion, the author presents his views on the general results which have been secured from the operation of these new trains and the probable trend in their future development.

DURING the last twenty years the travel habits of the American people have expanded greatly. The development of the low-priced automobile, and the construction of hard-surfaced roads have been the major factors in this change. In the face of increased travel, railway passenger traffic steadily declined until in 1933 it produced only about one third the gross revenue which it did in 1920. The data in Table 1 show this

TABLE 1 PASSENGER-TRAFFIC STATISTICS, CLASS 1 RAILROADS IN UNITED STATES

Year	Passenger revenue, dollars	Revenue per passenger mile, cents	Passengers carried
1921	1,153,792,000	3.09	1,035,496,000
1925	1,057,704,231	2.94	888,267,000
1930	729,470,279	2.72	703,598,000
1933	329,342,000	2.01	432,980,000
1934	346,325,993	1.92	449,775,000
1936	412,379,000	1.84	490,091,000
1937	442,809,000	1.79	497,288,000
1938	405,476,000	1.87	452,808,000

decrease both in the number of passengers carried and in the rates and revenues.

The large decrease in railway passenger traffic was undoubtedly due to the inroads of bus travel, private-automobile travel and, to a lesser degree, airplane travel. Prior to 1930, the railway managements, with properties still fairly prosperous because of good freight traffic, did little to remedy the situation except to reduce rates, as indicated in Table 1. They also eliminated some nonpaying trains and in the case of branch-line operation, where they were usually forced by governing bodies to continue passenger service regardless of losses involved, they substituted gasoline-electric motorcars for steam trains. These cars were definitely successful in reducing operating costs to such an extent that their first cost was paid for in a few years, but there was nothing

about their service to attract new passenger business or even to hold what was left. Their usage did influence later developments in that they brought to the attention of railway managements the possibilities of utilization of the internal-combustion motor in railway service. Their success in mechanical performance led the manufacturers to develop the new type of lightweight two-stroke-cycle Diesel motor, which is now playing such an important part in the operation of high-speed streamlined trains.

About 1930 the railroads started air-conditioning passenger cars. This was an instant success with passengers and though the cost was high, about $8000 per car, it became necessary to air-condition practically all main-line cars, with the result that there are now about 11,000 such cars in operation. In some of these cars there were other improvements made for the comfort of the passengers such as more comfortable seats and better lighting. The success of this development, from a traffic viewpoint, convinced the passenger-traffic officers that it might be possible to recapture at least a part of the lost business.

In 1900 the 60-ft wooden passenger car, seating about 80 passengers, weighed about 40 tons or 1000 lb per passenger. In the interests of safety, designs were changed first to include steel underframes, and in about 1910 to all-steel construction. Various parts were also increased in size and weight in order to secure greater strength and durability, until in the 1920's the 80-ft passenger car seating about 70 passengers weighed approximately 80 tons or 2300 lb per passenger. About 1930 still further weight was added, due to the introduction of air-conditioning equipment, weighing about 5 tons. As a result of this continued increase in weight of passenger cars, as well as a gradual increase in the speed of operation, the size of steam motive-power units steadily increased. Track structures and bridges had to be made stronger to carry these heavier locomotives and cars, and thus there was, from all of these angles, an increase in the cost of operating passenger trains, while at the same time the total traffic was diminishing and there was a continual demand for lower rates to save the remaining traffic.

DEVELOPMENT OF LIGHTWEIGHT CARS

About this same time some new structural materials became available for use by the car builders, namely, aluminum alloys, stainless steel, and various so-called low-alloy high-tensile steels. Because of these developments the railway engineers and the car manufacturers started active work in the design and construction of new trains of a lighter type for main-line operation. The first of these were small trains consisting of three or four cars with 600-hp Diesel engines located in the head car. They were made as light as possible in order to hold down the power requirements and to enable them to compete with buses and private automobiles by giving cheap transportation. The cars were of nonstandard height and some were narrower than the standard. They were streamlined to reduce power requirements and also to make an appeal to the public, already educated to such design by automobile developments. The reaction of the traveling public was so favorable that there was an immediate demand for more elaborate and larger trains for longer runs, with sleeping accommodations and other features for maximum comfort. Low first cost was soon lost sight of in this development. Interior decorators designed the most elaborate fittings and the trains were built with as many as 14 cars with power units totaling as

much as 5400 hp and auxiliary power for the cars of 1200 hp. A few of these trains were powered with streamlined steam locomotives but there was little change in the design of these locomotives from the old type except that higher driving wheels (84 in.) were used together with higher steam pressures (300 psi) and larger tenders to reduce delay in taking fuel and water. They also had roller bearings and some other refinements in design.

Most of the early trains had articulated cars because it was thought that this design reduced weight and gave better riding conditions. It developed however that neither of these factors proved to be important. There was a definite disadvantage in articulation particularly on long runs, in that standard cars could not be mixed with such cars. In case of difficulty, such as a bearing or wheel failure, a car failure might become a train failure. In the last few years practically all the lightweight cars have been built as individual units with standard floor and coupler heights and standard outside dimensions, so that they can be associated with any standard car. The weights of these cars were reduced about 40 per cent from those of the conventional structural-steel cars and thus four-wheel trucks could be used.

SPECIFICATION REQUIREMENTS FOR PASSENGER CARS

The only standards for design strength available at the time most of these cars were built were the provisions of the United States Railway Post Office mail-car specifications (1).[2] These specifications were first prepared in 1912 by a group of representatives of the Post Office Department, the Master Car Builders Association, and the car manufacturers. It was required that all postal cars meet these specifications. While other cars were not required to meet them, practically all passenger-train cars were designed at least to these requirements. In July, 1938, they were revised to clarify certain features. These new specifications were written so as to cover practically all types of construction with all types of materials. The major provisions of both the 1912 and 1938 specifications were, first, a minimum strength of center-sill construction based on 400,000 lb buffing static load and a factor of safety of two, and second, a minimum strength of end construction to protect against telescoping action. This latter required minimum section moduli for vertical end members and minimum shear values for their top and bottom connections. There were also numerous other limitations of stress in frame members.

Cars built to meet these specifications have rendered excellent service with a remarkably good record in the protection of passengers under wreck conditions. As a consequence when lightweight construction was started in about 1931 these specifications were used as a minimum requirement, even though there were no rules of the American Railway Association or of the Government setting such limitation. However, a feeling developed that because of the operation of lightweight cars between heavy cars in trains, it was necessary for the Association of American Railroads to set up more detailed specifications to cover the construction of all passenger-train cars built in the future, giving particular attention to prevention of damage from telescoping. A committee of the Association of American Railroads has recently developed these new specifications and they will probably govern all new construction of passenger cars. While these new specifications are generally based on the Railway Post Office specifications, they include some major changes: They provide for a center-sill strength such that an 800,000-lb load applied on line of draft will not produce any permanent deformation. The Railway Post Office specifications contained only a stress limitation and permitted the load application to be divided between the coupler line and the buffer line. This change will probably result in the

addition of more material than is now used in center sills of lightweight cars designed with high center sills and buffers, with little compensating benefit from added protection. The new specifications also include increased requirements for strength of couplers, carrier irons, and end construction, all of which mean some increase in weight, but which are apparently justified in view of the importance of maximum protection against telescoping of cars.

Generally speaking, the new lightweight cars are being designed and built just as strong and safe as the older heavy types. Better engineering, stronger materials, and improved methods of fabrication make this weight reduction possible. New welding technique largely eliminates the human element and when properly applied should be preferable to the old riveting practice.

NEED FOR LIGHT WEIGHT

Light weight is almost a necessity for high-speed operation in order to hold down the size and cost of power units, as well as the cost of operation. This is particularly true with Diesel-electric locomotives with their high first cost. Experience indicates that 450 lb per hp (including weight of locomotive) is a proper design ratio for high-speed operation. Some of the new trains have a power ratio as high as 550 lb per hp but the maintenance costs appear to run higher on such trains. In order to secure satisfactory performance on high-speed schedules, it is necessary to have some reserve power in the locomotive because some delays are inevitable and extra power is needed to make up the lost time. While the schedules may indicate that maximum speeds of only about 90 mph are required, actual operation requires at times speeds of 100 mph or even higher. This fact must be taken into consideration when designing locomotives and trains for this type of service.

CAR CONSTRUCTION MATERIALS

Several of the early trains were built of aluminum alloys but in recent years the majority of them have been built of stainless steel or low-alloy high-tensile steel. The major claims for the stainless steel are lower weight due to higher physical properties, better welding by the automatic shot-weld process, and high resistance to corrosion. Stainless steel does not have to be given paint protection of any kind, which reduces both first cost and maintenance cost. Its polished surface gives an attractive modernistic appearance to the cars and is easily cleaned. In high-speed operation the sandblast effect of particles from the ballast or right of way damages paint finishes quite rapidly. The low-alloy high-tensile steel advocates claim that they secure equally low weights for lower costs. Tests indicate that these steels have from two to four times the corrosion resistance of ordinary carbon steel. They are made to various formulas and have different trade names. It is claimed that some of them have somewhat better welding qualities than others.

Inasmuch as no exactly similar cars have been built using the different types of steel or aluminum, it is not possible to make accurate weight comparisons. While it is true that the full use of the superior physical properties of stainless steel or the light weight of aluminum is limited to some degree by deflection requirements, it should be possible, by using these materials instead of low-alloy steel, to construct a lighter car of equal strength. A general comparison of the large number of cars of the various types which have been built indicates a weight advantage for stainless steel and aluminum of at least 5 per cent. It would be expected that the construction cost of cars built of these materials would be somewhat higher due to the high cost per pound of such materials, but competition between builders has resulted in about equal bid prices.

It is too early to evaluate fully the different types of lightweight-car construction. Maintenance costs over a longer period

---

[2] Numbers in parentheses refer to the Bibliography.

TABLE 2 PHYSICAL PROPERTIES AND COMPOSITION OF MATERIALS
USED FOR PASSENGER-CAR BODIES

	Structural steel for cars	Typical low-alloy high-tensile steel	Stainless steel Grade 1	Stainless steel Grade 2	Typical aluminum alloy 17ST
Minimum yield point, 10^3 psi........	a	50	110	120	32
Minimum tensile strength, 10^3 psi....	b	70	140	150	50
Minimum elongation in 2 in., per cent	25	22	18	18	8 18
Modulus of elasticity, 10^6 psi........	28 30	28 30	27	27	10.5
Carbon, per cent....................	0.30	0.10	0.08 0.20	0.08 0.20	..
Manganese, per cent................	0.30 0.60	0.10 0.50	0.70 max	0.70 max	..
Copper, per cent....................	0.20	0.30	0	0	4.0
Chromium, per cent................	0	0.50 1.50	18	18	..
Nickel, per cent....................	0	0	8	8	..
Approximate cost, cents per lb.......	$2^1/_4$	$3^1/_2$	35	35	35

a One half tensile strength.   b 50,000 to 60,000 psi.

and performance in accidents will ultimately provide the answer. It appears, however, that a weight of about 100,000 lb for an 80-ft coach is entirely practicable with full safety to passengers. If too much space for extra luxuries is provided, part of the advantage of lightweight construction is lost, as the important factor is the weight per passenger carried. Table 3 shows the major parts of a typical lightweight car in which the weight saving has been made.

TABLE 3 COMPARATIVE WEIGHTS OF CONVENTIONAL AND TYPICAL LIGHTWEIGHT AIR-CONDITIONED PASSENGER COACHES

	All weights in pounds	
	Conventional steel coach	Lightweight stainless-steel coach
Length over end sills, ft..............	70	79
Seating capacity.....................	58	63
Weight of trucks.....................	52800	34000
Weight of car-body structure.........	66500	33000
Weight of inside furnishings..........	7000	5000
Weight of equipment.................	45400	25000
Total weight.................. ...	171700	97000
Weight per passenger................	2962	1540

RIDING CHARACTERISTICS OF LIGHTWEIGHT CARS

Due to higher operating speeds, the design of passenger-car trucks is a most important factor. The former six-wheel trucks were not entirely satisfactory at these increased speeds, and doubt existed in the minds of some engineers as to the possibility of securing good riding qualities with four-wheel trucks. The weights of the new car bodies were so low that only four axles were necessary to carry the load, affording further opportunity for weight reduction by the use of four-wheel trucks. Extensive experiments were made by various railroads and manufacturers, which resulted in the development of many new features of truck design with a decided improvement in riding characteristics.

Among these new features were the use of softer springs with greater deflections, hydraulic shock absorbers to control both lateral motion and vertical spring reactions, multiple-spring systems in triple-bolster trucks, and roll stabilizers. The use of cylindrical tread contours on wheels to reduce the nosing of trucks, and the grinding of treads to secure true rotundity and concentricity, were both helpful in softening the ride. Balancing of the wheels was also practiced to some extent, although a complete check of this feature has not been made due to lack of balancing machines in this country. It is now being given careful check on one railroad, which has installed the necessary machine. In an effort to reduce noise in the cars, rubber or other semiresilient

materials have been introduced in various parts of the trucks, such as pedestals, spring seats, and center plates. The mounting of the brake cylinders on the trucks instead of the body has also reduced the noise produced by the brakes. Improved insulation of car floors, as well as the sealing of windows, has also made a marked improvement, but there is still need for further noise elimination. The passengers can still hear the click of the wheels on the rail, the grinding of brake shoes on the wheels, and in some cases generator noise. Longer rails and better track maintenance would help to reduce rail and wheel noise and the use of disk-type brakes would reduce brake noise.

In some of the lightweight cars a certain amount of vertical body vibration was noticeable to passengers. This has been reduced by stiffening the car-body structure. The fore-and-aft vibration of cars which is so noticeable in many of the standard cars, particularly in operation with reciprocating steam locomotives, has been largely eliminated through the use of tight-lock couplers, which have little or no free slack and by softer acting draft gears.

The most unpleasant shocks in these cars are those of lateral acceleration, resulting either from nosing of trucks on tangent track, or the entering or leaving of curves. A novel attack on this problem is being made by one manufacturer, who has developed a pendulum type of car. This method of mounting the car on the truck differs completely from standard practice, in that the car body is virtually suspended from the truck, floating on soft vertical helical springs located well above the center of gravity of the car body. These springs permit through horizontal deflection the necessary truck motion relative to the body, and this motion is positioned and controlled by a pair of horizontal links, which are restrained by rubber disks and shock absorbers acting between the car body and vertical extension of the truck frame at a point above the center of gravity of the body. In going around curves this type of car tends to bank and thus avoids the unpleasant sensation resulting from car-body roll. Preliminary tests of this type of car indicate an improvement in riding characteristics particularly on curves.

In general it appears that there is still need for research in truck design, in order to secure even greater comfort for passengers and to make possible higher operating speeds on sharp curves which, particularly on western roads with much track curvature, are a definitely limiting factor to fast over-all operating schedules.

Improvements in track structure and maintenance are also necessary if still higher speeds are to be attained. Whatever de-

sign of truck is used, satisfactory riding characteristics cannot be secured in high-speed operation over bad track.

BRAKING

The introduction of high-speed lightweight trains brought many new problems in braking. The stopping of these trains within reasonable distance is a difficult problem. While the car weights are less, individual wheel loads are as high and in some cases (particularly with the articulated type) higher than with the old heavy cars. In order to reduce the stopping distance, there have been a number of changes from former practice, such as truck mounting of brake cylinders which increases rigging efficiency, much higher braking ratios (some as high as 250 per cent), decelakrons to reduce pressures as the train slows down and thus prevent sliding of wheels, automatic car sanders acting only in emergency application, and electropneumatic control. In some cases the very high brake ratios resulted in considerable wheel and brake-shoe trouble. With all of these improvements there is still a question whether emergency stopping distances are as short as they should be. To effect shorter stops without undue punishment to wheels and shoes, new types of disk brakes are being tested. One train so equipped is now in operation. This type of brake has the advantage of a more uniform coefficient of friction throughout the speed range, thus permitting brake pressures which more closely approach the coefficient of adhesion between the rail and the wheel. Unfortunately, this coefficient of adhesion varies widely under different rail conditions. Thus if higher braking pressures are used, there is danger of sliding under unfavorable conditions. To meet this problem, either car sanding equipment or governors to reduce pressures automatically, when sliding begins, may prove to be a necessary complication.

In both steam- and Diesel-powered lightweight trains, the locomotive weight is a large percentage of the train weight, but in most cases locomotives do not do their full share of the braking. The steam locomotive is particularly lacking for three reasons: First, there are often some unbraked wheels on engine trucks for operating safety; second, the large spread between light and loaded weight of tenders makes it necessary to design braking power on the basis of light weight, and third, enginemen have to relieve engine brake pressure frequently to guard against the loosening of driving tires due to heating. The Diesel is better in that all wheels are braked and the spread between light and loaded weight is relatively small, but the small-diameter wheels are so highly loaded that the braking ratio must be limited in order to protect the wheels from damage. It would be advantageous if Diesel-electric locomotives had a supplementary dynamic electric brake such as is used in some electric locomotives as well as in the new experimental turboelectric design.

COMPARISON OF DIESEL-ELECTRIC AND STEAM LOCOMOTIVES

It may be noted from Table 7 that the Diesel-electric locomotive is more commonly used in these new high-speed streamlined trains than is the steam locomotive. It is claimed by some that this is chiefly due to the interest of the public in a new type of power, but the users apparently have good engineering arguments to support their choice. The advantages claimed for the Diesel are: First, high availability; second, rapid acceleration because of high tractive power at lower speeds; third, low maintenance costs; fourth, low fuel cost (usually less than half that of steam); fifth, lower rail stresses. The steam locomotive is less expensive, costing about $37 per hp. However, it is harder on track at high speeds due to the dynamic augment produced by overbalance in the driving wheels. It appears that a radical change in the design of the steam locomotive is necessary to make it capable of fully competing with the Diesel in this service. The important objects of such redesign should be the securing of a more con-

stant torque and the reduction of track punishment. Three outstanding locomotives of new design are under construction or in service, namely, the 5000-hp turboelectric being manufactured by the General Electric Company for the Union Pacific Railroad, the Baltimore and Ohio Railroad's four-cylinder engine, and the Pennsylvania Lines' four-cylinder high-speed engine. All of these new types are experimental, but they indicate a trend toward a type of steam locomotive having more constant torque. A step still further in this direction is involved in a new Baltimore and Ohio design, which includes individual axle drive.

The comparison of typical modern high-speed steam locomotives and Diesel-electric locomotives on a horsepower basis, as shown in Table 4, is open to criticism because of the basic differ-

TABLE 4 COMPARISON OF TYPICAL MODERN HIGH-SPEED STEAM AND DIESEL-ELECTRIC LOCOMOTIVES

	Diesel-electric	Steam 4-6-4	Steam 4-8-4
Nominal hp of Diesel motors	4000	....	....
Maximum indicated hp	....	4300	5750
Weight with fully loaded tender, lb.	608000	809000	896000
Weight, lb per hp	152	188	156
Approximate cost per hp	$87.50	$35.00	$34.00

ence in the horsepower curves of the two types and also the difference in method of power transmission. The nominal horsepower of the Diesel-electric locomotive is available over practically the entire speed range, whereas the maximum indicated horsepower of the steam locomotive is available only during a limited speed range. With locomotives of equal horsepower, the Diesel has the advantage of greater power for accelerating at low speeds and greater sustained horsepower at very high speeds. Some difficulties with pistons, heads, and liners have been experienced in services where Diesel-electric locomotives were operated with such loads and schedules as to require almost continuous maximum output of power. Recent improvements in materials and designs of these parts have largely overcome this trouble. Nevertheless, the maintenance cost of any type of locomotive increases as the working-load factor increases. Proper design of a motive-power unit should provide some reserve power for emergency use, but not such an excess as to result in uneconomically high cost of construction.
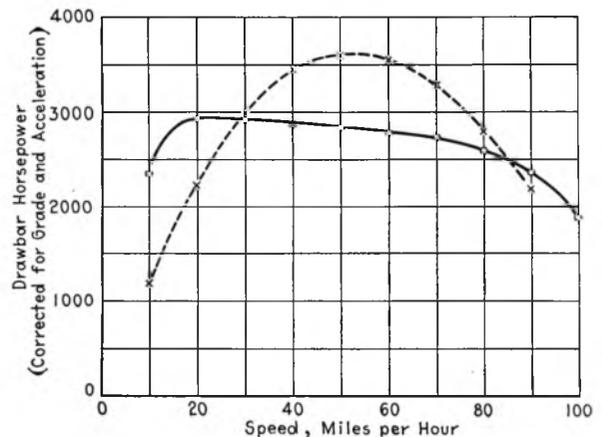
FIG. 1 COMPARATIVE DRAWBAR-HORSEPOWER CURVES FOR STEAM AND DIESEL-ELECTRIC LOCOMOTIVES
(Solid curve. Diesel-electric locomotive, 4000 nominal hp. Dotted curve, 4-6-4 steam locomotive, 4300 max ihp.)

Fig. 1 shows the relative drawbar horsepower of a 4000-hp Diesel-electric locomotive and a modern (1938) high-speed steam locomotive of the 4-6-4 type with 4300 max ihp. These curves are plotted from actual dynamometer-car readings with necessary corrections for grade and acceleration. They show the advantage

of the Diesel-electric locomotive at low speed and very high speed and also the advantage of the steam locomotive at intermediate speeds, such as would exist in ordinary grade working. This advantage in grade climbing is offset to a considerable extent by the 33 per cent greater weight of the steam locomotive.

The sharp dropping off of the horsepower curve for steam locomotives at high speed is due largely to low mean effective pressure in cylinders. To improve this condition, radical changes both in the design of valve and of valve gears appear to be necessary.

In the case of Diesel-electric locomotives, the curve does not drop off as sharply as does that for steam locomotives and further improvement has already been secured in the latest designs by change in the generator voltage control. This change has also increased the power output at low speeds.

In regular operation of the two types the performance of the Diesel-electric locomotive is usually that which is indicated by the drawbar-horsepower curve, but in the case of the steam locomotive there are a number of variables which may result in average performance somewhat below what is indicated by the horsepower curve derived from tests. The skill of the engineer varies and the condition of the locomotive is not always as nearly perfect as it is in the test.

Manufacturers of Diesel-electric locomotives have an advantage in that they produce a practically standardized product and thus can reduce cost through line production. On the other hand almost no two lots of steam locomotives are the same. Each order means a new design with resulting high costs for engineering and production. This also means higher cost for and larger stocks of repair parts. American railroads have made a marked advance in the standardization of freight cars but there has always been strong opposition to any plan for standardization of steam locomotives or passenger cars, due mainly to differences of opinion among different managements. The usual explanation is that operating conditions vary but this is hardly a valid excuse at least when certain railroads are compared.

The use of Diesel-electric locomotives is more extensive on western roads than on those in the East. This is probably due to the following factors: Most of the oil production is in this territory, which results in lower prices, and the traffic in oil and its by-products is a more important source of revenue than that from coal, whereas the reverse is true in the eastern territory. In the Southwest most of the steam locomotives use oil as fuel because of its lower first cost. The usual Diesel-locomotive fuel is about 28-gravity furnace oil costing about four cents per gallon at the refinery. Steam-locomotive fuel oil is a low-grade residuum costing about two cents per gallon at the refinery. The boiler-feedwater supply and treatment is also much more of an expense and problem in at least a large part of the West than it is in the East.

Diesel-electric locomotives provide better vision for the engineer because the cabs are located at the head end. The riding quality of Diesels is better than that of steam locomotives. They also provide smoother-riding trains as their high tractive effort at low speeds enable them to start and accelerate trains more quickly and smoothly.

Steam locomotives have an advantage in that more horsepower can be put in one unit. It may be noted in Table 4 that the increase in horsepower is secured with a lower percentage of increase in first cost and weight, that is, both the cost and weight per horsepower are lower for the 4-8-4 locomotives than for the 4-6-4 type. The maintenance cost per horsepower should also be somewhat lower for the 4-8-4 locomotives. It is true that any number of Diesel units can be coupled and controlled from one cab, but this increase in number of units means increased first cost and increased weight. A 2500-hp Diesel unit would appear

to be a desirable size for heavy service, particularly in mountainous territory, but this does not appear practicable at present.

The performance records of Diesels show fewer failures than similar records for steam locomotives, even when the latter are operated in slower-speed service. The multiunit power plant is a helpful feature in the prevention of failures. If something goes wrong in one power plant the locomotive can still go on to its terminal and in many cases the defect can be repaired en route with little or no delay. A major breakage or defect on a steam locomotive means a failure.

If an entire territory were completely Dieselized, major savings could be secured in elimination of water and fuel facilities and also intermediate terminals. With partial Dieselization, it is necessary to provide certain special facilities for water and fuel, but these are not expensive. Diesel-electric locomotives are designed to operate about 700 miles without refueling or watering, whereas steam locomotives with large tenders operate only about one half of this distance for fuel and one third the distance for water. Fewer and shorter stops for fuel and water are essential in making high-speed schedules.

Diesel-electric locomotives have a higher availability than steam locomotives because they need less servicing. Furthermore, individual major parts can be changed in a short time, thus avoiding delay of the unit for repairs. In service they are regularly making $1\frac{1}{2}$ to 2 times as much mileage as steam locomotives, though it is only fair to say that part of this may be due to their use on preferred runs. Diesel-electric locomotives are cleaner in their operation, due to lack of smoke. This is particularly advantageous in cities and towns. It also adds to passenger comfort in that the view from the car windows is not obstructed by smoke.

Table 5 gives a breakdown of the weight of a typical 2000-hp

TABLE 5 WEIGHT OF STANDARD 2000-HP DIESEL-ELECTRIC LOCOMOTIVE

	Per cent
Two 1000-hp Diesel motors and auxiliary equipment	23
Two electric transmissions with auxiliary equipment	21
One steam boiler with auxiliary equipment	2
Two trucks (less traction motors)	28
One car body	26
Total	100

Total light weight	277000 lb
Weight of oil (1420 gal)	11000 lb
Weight of water (1300 gal)	10800 lb
Weight of sand (16 cu ft)	1600 lb
Total	300400 lb

Diesel-electric locomotive. The large percentage of the weight involved in the trucks may be noted. This is due to the heavy construction needed to support the traction motors and withstand the shocks of high-speed operation. The trucks are constructed of alloy steel and there appears to be little opportunity of making them lighter. The Diesel motors are almost as light as they can be made to withstand the service, but it may be possible in future development to take more power out of each cylinder, which would result in a reduction of weight per horsepower. The transmission accounts for a considerable percentage of the total weight. There is some possibility of reducing both the cost and weight of this part through the development of suitable hydraulic transmissions, but no large locomotives of this type have as yet been built in this country. The weight of the body structure cannot be reduced to any great extent as alloy steel and welded construction are already used throughout.

The center of gravity of the Diesel-electric locomotive is considerably lower than that of steam locomotives and this fact, together with the shorter rigid wheel base makes it possible to operate the Diesel-electric locomotive at about 15 per cent higher speeds on sharp curves. This is an important factor in making high-speed schedules, particularly on runs where there is a great deal of curved track.
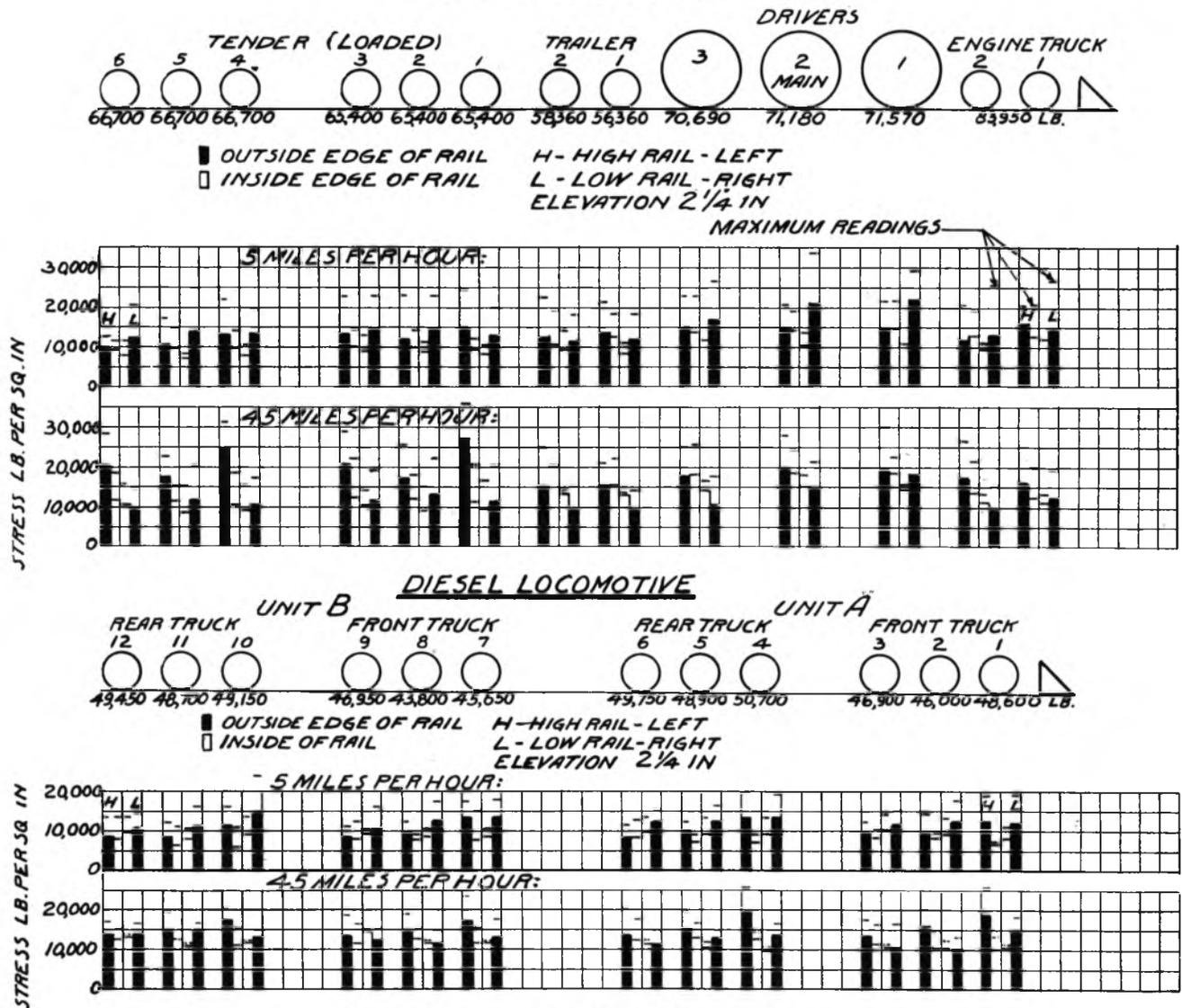
FIG. 2 MEAN STRESSES IN RAIL BASE UNDER EACH WHEEL
(Five-deg curve, 110-lb rail.)

RELATIVE COST OF MAINTENANCE OF DIESEL-ELECTRIC AND
STEAM LOCOMOTIVES

There is a great deal of controversy as to the relative cost of maintenance of Diesel-electric locomotives and steam locomotives. It is difficult to secure reliable data from which to make comparisons. The comparison of costs on different railroads is open to question because of difference in operating conditions and accounting practices. The two types of power are not regularly used in the same type of service on the same railroad. Furthermore, the Diesel-electric locomotives are rather new and very few of them have received general repairs. It may be said that the two types are definitely different in regard to general repairs; the steam locomotive because of its boiler, flues, and firebox must be given general overhauling at relatively frequent periods. The Diesel-electric locomotive, on the other hand, is being constantly repaired by the regular replacement of the wearing parts. Some roads have put them in the back shop for general repairs after about 750,000 miles of service. It was originally thought that this might be the life of the major element of the Diesel motor, the crankshaft, but it has developed that the life of this part will normally be greatly in excess of this figure, and some roads are planning on one million or more miles between general shoppings.

The best available data appear to be the five-year-period maintenance-expense figures of a group of Pacific-type locomotives with approximately 3800 max ihp capacity, operating in normal passenger service, and similar figures for a group of 3600-hp Diesel-electric locomotives in high-speed service on the same railroad. These data show that the steam locomotives cost about 20 cents per mile for maintenance, whereas the Diesel-electric locomotives cost approximately 17 cents per mile for maintenance. These particular Diesels have never been given a general overhauling and, therefore, in order to be conservative, it might be fair to add as much as 3 cents per mile to this figure. These data indicate that it may be conservatively claimed that Diesel-electric locomotives can be maintained for at least the same cost as steam locomotives. If this same group of steam locomotives were operated at high speed, the repair costs would
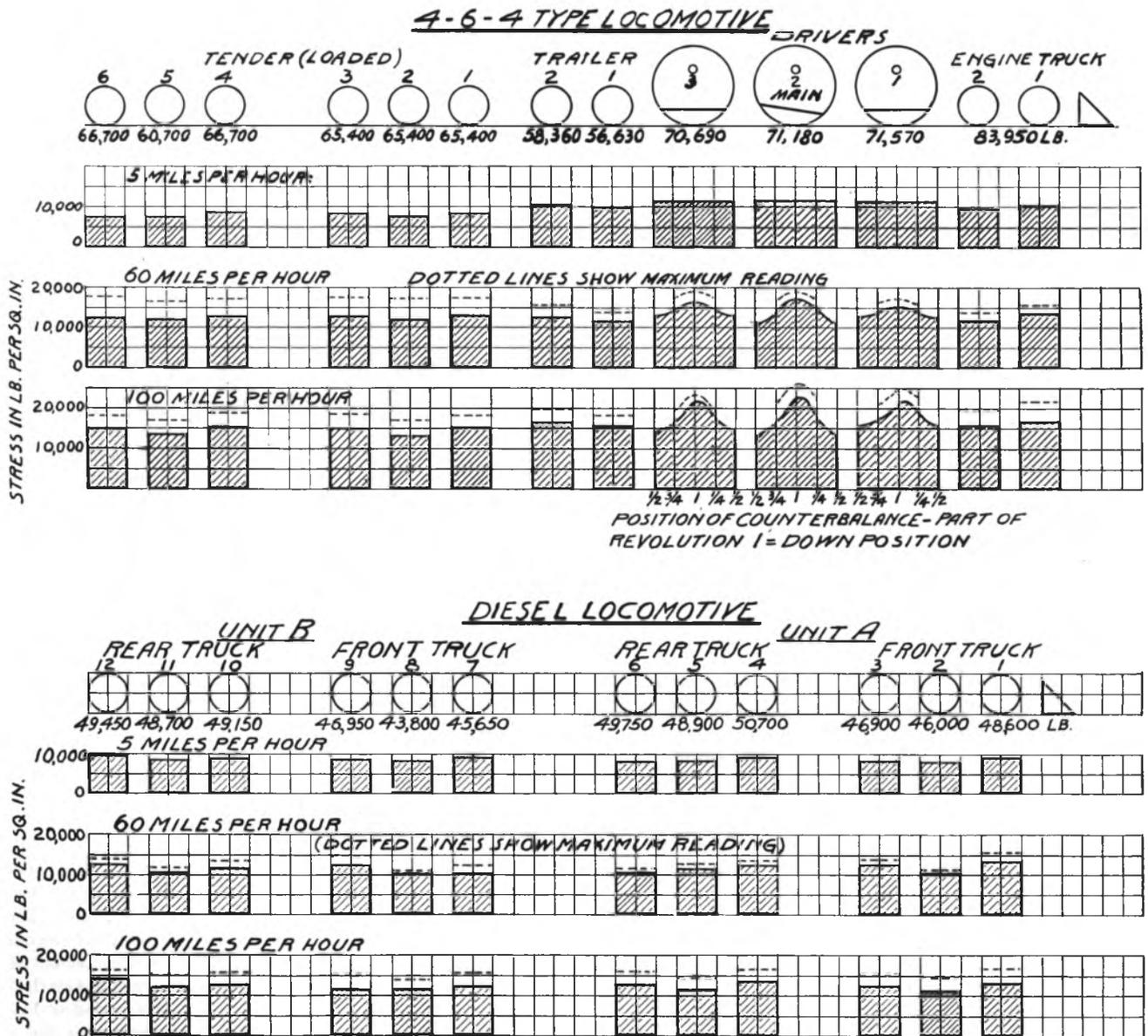
FIG. 3 VERTICAL STRESSES IN BASES OF BOTH RAILS
(Tangent track, 110-lb rail.)

undoubtedly be greater than the 20 cents shown. It is claimed by the builders of steam locomotives that the most modern type will have a lower cost of maintenance than older types, but it can also be claimed that the details of Diesel-electric locomotives are being steadily improved with a resultant reduction in maintenance cost. It must be recognized that costs on different railroads vary widely, particularly those applying to Diesels, since they are new and maintenance forces are not as familiar with them as they are with the steam locomotive. Furthermore, most railroads' repair facilities were designed for steam-locomotive work and are not suitable for efficient handling of Diesel repair work. New facilities are gradually being provided for taking care of Diesel work and their use should result in some reduction in maintenance cost.

TRACK STRESSES

Every steam locomotive must have a certain amount of overbalance in the driving wheels, in order to balance the reciprocat-

ing parts. This overbalance results in a dynamic augment which increases as the square of the speed and thus becomes an important factor in producing stresses in the rail and roadbed at high speed. In some cases this effect has been so serious that the main drivers actually lift from the rail at high speed and may produce kinked rail. Improvement in this condition has been made by reducing the weight of the reciprocating parts to a minimum and reducing the percentage of the reciprocating weight balanced to a minimum limited by the production of nosing and fore-and-aft motion of the engine. A still further effort along this line is embodied in the new experimental designs of locomotives having two complete sets of cylinders and motion work. This not only reduces the weight of reciprocating parts but gives another desirable characteristic, that of more constant torque. In order to illustrate the difference in rail stresses produced by a large Diesel-electric locomotive which has no reciprocating parts and a large modern high-speed reciprocating steam locomotive, Figs. 2 and 3 are included. These show the results of

rail-stress measurements on 110-lb rail with a modern 4-6-4 steam locomotive and a 3600-hp Diesel-electric locomotive. Fig. 2 shows stresses on a 5-deg curve at speeds of 5 and 45 mph and Fig. 3 stresses on level tangent track at speeds of 5, 60, and 100 mph. In Fig. 2 the stresses are shown separately for both edges of both rails, and the maximum stress at each location, as represented by the average of three highest readings at each location, is shown in short horizontal lines. In Fig. 3 the average vertical stress in the rail base is shown and the effect of the counterbalance by indicating its relative position. The maximum stress is also an average of three maximum values. The damaging effect of counterbalance is clearly shown in these data which are actual stress measurements and not calculations.

If a steam locomotive slips at high speed, the rotational speed of the driving wheels may be greatly in excess of that involved in these tests, therefore, the dynamic augment and resulting rail stress are greatly increased. The high rail stresses under steam-locomotive tenders due to too-high loads on small-diameter wheels may also be noted. Large tenders are necessary on high-speed locomotives in order to reduce to a minimum the delays for taking fuel and water, but it appears necessary to add two more pairs of wheels or increase the diameter of the wheels, in order to reduce wheel loads to a reasonable figure where damage to both track and wheels is not excessive.

No complete data are available for rail stresses under 4-8-4 type high-speed steam locomotives. However, the tests made indicate that rail stresses under the main drivers of such locomotives are about 25 per cent greater than those shown for the 4-6-4 type. This is due to the fact that the reciprocating parts in these larger and higher-powered engines are heavier and the driving-wheel diameter is usually made smaller to meet heavy-grade operating conditions.

All of these track-stress data point to the need of a radical change in steam-locomotive design for competition with the Diesel-electric locomotive in high-speed passenger-train operation. The four-cylinder design is a step in the right direction. The individual axle drive proposed by the Baltimore and Ohio Railroad is a further step and also the new turboelectric type. Up to the present time the gear-drive turbine locomotive has not been tried out in this country, but it too may have possibilities. The individual axle drive for a steam locomotive appears to have the most interesting possibilities since it involves not only relatively constant torque but also better track protection resulting from complete balance. A lower center of gravity can also be secured since driving wheels of smaller diameter can be used. A special committee appointed by the Association of American Railroads composed of leading railway mechanical engineers and representative engineers of the locomotive builders is making a thorough study of the entire question of design of steam locomotives for high-speed operation. Out of their work there will probably come some concrete and valuable recommendations on this important question.

Until some marked changes are made in steam-locomotive design it appears that the Diesel-electric locomotive will have a definite advantage in so far as effects on rail and roadbed are concerned.

Lightweight passenger cars also have an advantage over conventional cars in the matter of track-maintenance costs. Though their individual wheel loads are just as high, their lower total weight and trucks with shorter wheel base result in lower lateral rail pressures in high-speed operation.

Thermal Efficiencies

The theoretical maximum thermal efficiency of the Diesel-electric locomotive figured on the basis of energy developed at the rail is about 27 per cent, whereas the corresponding figure for a modern steam locomotive seldom exceeds 7 per cent, which would indicate a relative fuel consumption of about 1 to 4. This is somewhat altered for over-all performance by difference in weight of the two types of locomotive and difference in stand-by losses. It is difficult to find accurate comparative fuel-consumption data because the two types of power are not ordinarily used on similar runs. One comparison available is on a main-line fairly high-speed through train where steam power was superseded by Diesel-electric locomotives. The relative round-trip fuel-consumption figures based on a 12-month average on this run are as follows: Steam locomotives used 124 tons of coal; Diesel-electric locomotives used 3950 gal of Diesel oil. Based on the Btu content of the two types of fuel, these data indicate that the thermal efficiency of the Diesel-electric locomotives figured at the drawbar was 5.5 times that of the steam locomotives.

A careful check was made on another railroad to determine the relative amount of Diesel oil consumed by 3600 hp Diesel-electric locomotives and modern Pacific-type oil-burning steam locomotives, when pulling both light and heavy medium-speed main-line passenger trains on a 550-mile run. These data show the steam-locomotive oil consumption to be approximately four and one half times that of the Diesel-electric locomotives.

There appears to be little chance of improving the efficiency of the Diesel-electric locomotive except in the factor of transmission loss. There should be a possibility of marked improvement in steam-engine efficiency because of the many losses now involved. There has been a steady improvement in steam-locomotive thermal efficiency as indicated by the figures of the Association of American Railroads, which show that in 1920 it took 173 lb of coal to produce 1000 gross-ton-miles of freight transportation and 18.8 lb of coal per passenger-car-mile, whereas in 1938 these figures were only 113 lb for freight and 14.7 lb for passenger trains. A large part of this fine showing is undoubtedly due to improvements in steam-locomotive design and the reductions would be even greater if all the locomotives in service were of a modern type. Credit is due the designers for these improvements, which incidentally have been developed without the urge of competition from other types of power. Naturally under noncompetitive conditions the designers have been more or less bound by precedent. Conditions have now been changed. The young virile Diesel locomotive (whose designers were not swayed by precedent and who had at their disposal the vast amount of research data developed by the automotive industry) has reached a remarkable state of development in the short period of five years. This new competition should prove to be a spur and a benefit to the steam-locomotive designers, as is indicated by the numerous new designs which are mentioned elsewhere in this paper.

The Diesel designers had an advantage in connection with high-speed operation in that they started their development of road locomotives on the basis of high speed (top speed 118 mph). On the other hand, steam-locomotive designers focused their attention on the improvement of thermal efficiency and greater power development with top speeds of about 90 mph. Thus when the demand for high-speed operation developed so rapidly, their designs were not entirely adequate. This feature of high speed is the one which is now being given major consideration, and improvement should result.

Electric Locomotives

The modern electric locomotive is also a highly satisfactory type of motive power for high-speed operation, both from the viewpoint of capacity and effect on track structure. Early types had detrimental effects on track due to high lateral forces, but later designs have overcome this. The use of this type of power in this country has been limited, due to the fact that the

density of traffic on most railroads does not warrant the high cost of overhead construction. Where it is used for long-distance high-speed operation, as on the Pennsylvania Railroad, between New York and Washington and Harrisburg, it has shown excellent performance, but there appears to be little probability of electrification of less dense traffic lines, particularly those in the western part of the United States.

OPERATING EXPENSES

In 1938 Coverdale and Colpitts (2) issued a comprehensive report covering the results of their studies of operating costs and revenues of a majority of the streamlined trains in operation at that time. The cost data for the various railroads show such wide variations that it is difficult to arrive at a fair average figure for cost of operation of Diesel locomotives as compared with steam. Unfortunately, there are no roads on which comparable trains are operated with steam and Diesel-electric locomotives, so that there is no fair basis for comparison. However, a study of all the data in this report indicates that the cost of maintenance of a 3600-hp Diesel-electric locomotive is no greater and probably less than that of a modern steam locomotive of similar power, handling equal-weight trains on fairly comparable schedules. The fuel cost is approximately one half as much. The cost of crew wages is about the same, though it should be noted that on some roads a maintainer is included in the crew of Diesels for long-distance operation. This should not be necessary when firemen are fully trained in their work on this type of power. Lubrication costs (a lesser factor) are greater for Diesels than for steam locomotives. It is noticeable that the maintenance costs for Diesels are higher when the total load is much above 450 lb per hp.

The outstanding feature of this report is the large margin between gross revenue and net revenue (depreciation charges are not included). Most of the trains show net revenue over 50 per cent of gross revenue and in some cases this is as high as 75 per cent. These figures, are, of course, unduly favorable because of the omission of two important factors, namely, the overhead charges for depreciation and interest, and the item of track maintenance. The overhead in one particular case is shown in Table 6 of this paper. Even with the high mileage made by such trains, the overhead is a major item of expense because the first cost of these trains is high. As to track maintenance, it is well recognized that safe and comfortable high-speed operation necessitates a high standard of track maintenance. Furthermore high-speed operation, particularly with steam locomotives, increases the cost of maintenance, though no definite figures are available to determine such extra cost. It is a fact that large expenditures have been made, particularly on the western railroads to eliminate curvature and alter curves to make high-speed operation successful. As an example of this, it may be noted that one transcontinental railroad spent over four million dollars on such work during the last three years, and similar work is still under way. The total cost of these improvements should not be charged against the operation of the new trains for these changes also benefit the operation of other trains. Nevertheless, it must be admitted that there is a considerable part of this extra expense, which should properly be charged against the new trains to give a full picture of the financial results. If speeds are still further increased, this factor will become of even greater importance as more work of this kind will be necessary to make such schedules.

COACH-TYPE STREAMLINERS ON LONG RUNS

Table 6 shows a comparison of weights and operating costs between two streamlined trains operating in similar transcontinental service on one railroad. One of these trains is a first-class

all-sleeping-car train and the other is an all-coach train. All of the data are prepared on the basis of full load of passengers which,

TABLE 6 COMPARISON OF OPERATING COSTS FOR A FIRST-CLASS ALL-SLEEPING-CAR STREAMLINER AND AN ALL-COACH STREAMLINER

Item	First-class train	All-coach train	Estimated 10-car all-coach train
Number of cars...................	9	5	10
Number of Diesel power units......	2	1	2
Horsepower of locomotive..........	3600	1800	3600
Total weight of cars, tons..........	482	248	496
Weight of power units, tons........	287	147	287
Total weight of train, tons..........	769	395	783
Length of train, ft.................	890	473	970
Total salable seats or berths........	121	156	312
Out-of-pocket cost per mile of operation[a]........................	$0.88	$0.68	$0.88
Train length per passenger (full load), ft...............................	7.4	3.0	3.0
Car weight per passenger, tons......	4.0	1.6	1.6
Train weight per passenger, tons....	6.4	2.5	2.5
Out-of-pocket cost per passenger per mile of operation................	$0.0073	$0.0043	$0.0028
Total yearly operation, miles.......	232,000	232,000	232,000
Total first cost of train.............	$1,110,000	$580,000	$1,160,000
Overhead cost per mile, based on 10 per cent of first cost.............	$0.48	$0.25	$0.50
Overhead cost per passenger per mile (full load)......................	$0.004	$0.0016	$0.0016
Overhead and operating cost per passenger-mile (full load)............	$0.0113	$0.0059	$0.0047

[a] This cost does not include sleeping-car employees' wages or maintenance expense for sleeping cars.

of course, is not always the case. Estimated data are also shown for a train similar to the coach streamliner but with twice the capacity. All operating costs are taken from the Coverdale and Colpitts report (2).

It may be noted that the train weight per passenger for the first-class train is 6.4 tons, whereas for the coach train it is only 2.5 tons, which accounts largely for the higher out-of-pocket cost for the former. The overhead cost per passenger is more than twice as great. The total overhead and operating cost per passenger-mile is approximately twice as great for the first-class train as for the coach train. It is thus apparent that much lower rates can be charged on coach trains. This latter type of train has been remarkably successful in attracting passengers for this long run, although it involves their spending two nights on the road.

It is the coach type of train that primarily has the possibility of diverting traffic from private automobiles and buses to the railroads. Because of the low rates it also has the possibility of developing a large amount of new traffic among people in the low- and medium-income group, who are becoming more and more travel-minded and have more and more time available for such travel. There is almost no limit to the possibilities for development of this type of travel. People of this class want low-cost transportation with high speed and comfort. All of these are being provided in such trains.

The Coverdale and Colpitts report shows remarkably successful results from the operation of all of the high-speed streamlined trains. They have undoubtedly brought back to the railroads some of the lost traffic and give promise of developing new traffic.

It appears that these new high-speed trains, particularly those in the western territory are now operating at about the maximum speed, which is possible on the present track structure. The limitation to further increase in speed lies in the large number of rather sharp curves which are unsafe for too high speed and uncomfortable for passengers when taken too fast. The number of unprotected grade crossings is also a limitation, as well as speed restrictions through towns with grade crossings. It would be very expensive to correct all of these track conditions.

Outside of the passenger-traffic gain, these new trains have been warranted by the public interest which they have aroused.

TABLE 7   LIST OF HIGH-SPEED LIGHTWEIGHT STREAMLINED PASSENGER TRAINS

Railroad	Name of train	No. of trains	Service	Type of loco-motive	Run	Route, miles	No. of cars
A. T. & S. F.	Super-Chief	2	Twice weekly	Diesel 3600 hp	Chicago Los Angeles	2229	9
A. T. & S. F.	The Chief	6	Daily	Steam	Chicago Los Angeles	2229	10
A. T. & S. F.	El Capitan	2	Twice weekly	Diesel 1800 hp	Chicago Los Angeles	2229	5
A. T. & S. F.	San Diegan	1	Twice daily	Diesel 1800 hp	Los Angeles San Diego	126	7
A. T. & S. F.	Chicagoan Kansas Cityan	2	Daily	Diesel 1800 hp	Chicago Wichita	678	7
A. T. & S. F.	Golden Gate	2	Twice daily	Diesel 1800 hp	Bakersfield San Francisco	316	6
Baltimore & Ohio (Alton)	Abraham Lincoln	1	Daily	Diesel 1800 hp	Chicago St. Louis	282	8
Baltimore & Ohio (Alton)	Ann Rutledge	1	Daily	Steam	Chicago St. Louis	282	8
Boston & Maine	Flying Yankee	1	Daily ex. Sun.	Diesel 600 hp	Boston Portland	114	3
C. B. & Q.	Pioneer Zephyr	1	Daily	Diesel 600 hp	Kansas City Omaha-Lincoln	250	4
C. B. & Q.	Sam Houston Texas Rocket	2	Twice daily	Diesel 600 hp	Fort Worth Dallas-Houston	283	4
C. B. & Q.	Mark Twain	1	Daily	Diesel 600 hp	St. Louis Kansas City	279	3
C. B. & Q.	General Pershing	1	Daily	Diesel 1000 hp	St. Louis Kansas City	279	4
C. B. & Q.	Denver Zephyrs	2	Daily	Diesel 3000 hp	Chicago Denver	1036	12
C. B. & Q.	Twin City Zephyrs	2	Twice daily	Diesel 1800 hp	Chicago Twin Cities	419	8
C. M. & St. P.	Hiawatha	4	Twice daily	Steam	Chicago Twin Cities	422	9
C. R. I. & P.	Rocket	1	Twice daily	Diesel 1200 hp	Chicago Peoria	161	4
C. R. I. & P.	Rocket	1	Daily	Diesel 1200 hp	Chicago Des Moines	358	4
C. R. I. & P.	Rocket	2	Daily	Diesel 1200 hp	Kansas City Minneapolis	489	3
C. R. I. & P.	Rocket	2	Daily	Diesel 1200 hp	Kansas City Dallas	627	4
Gulf, Mobile & Northern	Rebel	2	Daily	Diesel 660 hp	New Orleans Jackson	488	4
Gulf, Mobile & Northern	Rebel	1	Daily	Diesel 660 hp	Mobile Union	181	1
Illinois Central	Green Diamond	1	Daily	Diesel 1200 hp	Chicago St. Louis	294	4
New York Central	Mercury	1	Daily	Steam	Cleveland Detroit	165	4
New York Central	Twentieth Century	2	Daily	Electric & steam	New York Chicago	960	9+
N. Y. N. H. & H.	Comet	1	6 trips daily	Diesel 400 hp	Providence Boston	44	3
Pennsylvania	Broadway Limited	2	Daily	Electric & steam	New York Chicago	907	8+
Reading	Crusader	1	Twice daily	Steam	Philadelphia Jersey City	90	5
Seaboard Air Line	Silver Meteor	1	Twice weekly	Diesel 2000 hp	N. Y.-Miami N. Y.-St. P.	1389 1247	7
Southern Pacific	Sunbeam	1	Daily	Steam	Houston Dallas	264	8
Southern Pacific	Daylight	2	Daily	Steam	Los Angeles San Francisco	471	14
Union Pacific	City of Salina	1	Daily	Diesel 600 hp	K. C.-Salina K. C.-Topeka	187 68	3
Union Pacific	City of Portland	1	5 trips monthly	Diesel 1200 hp	Chicago Portland	2272	6
Union Pacific	City of Los Angeles	1	5 trips monthly	Diesel 2400 hp	Chicago Los Angeles	2298	13
Union Pacific	City of Los Angeles-2	1	5 trips monthly	Diesel 5400 hp	Chicago Los Angeles	2298	17
Union Pacific	City of Denver	2	Daily	Diesel 2400 hp	Chicago Denver	1048	10
Union Pacific	City of San Francisco	2	5 trips monthly	Diesel 5400 hp	Chicago San Francisco	2259	14

NOTE: Six more trains for various railroads are under construction at the time of preparation of this tabulation, and also a considerable number of lightweight cars for use in other trains.

Prior to their advent, there was a feeling that the railroads had not kept pace with developments as they should have, and consequently the public had less interest in their welfare. They are now much more favorably inclined and this may have an influence on future legislation.

Large numbers of new lightweight cars are now being added to main-line trains, and there is a gradual replacement of old types. The public demand for such cars is due largely to their many comforts and attractive appearance. There is every indication that there is a real revolution in American passenger trains. Table 7 shows a list of the high-speed streamlined trains now being operated on the railroads of the United States.

CONCLUSIONS

The streamlined high-speed passenger train has been definitely successful on American railroads in recovery of traffic, financial return, and mechanical performance. The number of such trains in service should continue to increase.

Both steam and Diesel-electric locomotives will pull these new trains, but the Diesels will predominate unless new designs of steam locomotives prove themselves better suited for the long high-speed runs than are the present types. There should be a continual improvement in the design details of Diesel-electric locomotives. Radically new designs of steam locomotives will probably be built to meet the high-speed operating conditions.

Schedules faster than those now being made are possible and probable as soon as roadway conditions are further improved by curvature reduction and grade-crossing elimination.

BIBLIOGRAPHY

1 "Railway Post Office Specifications for the Construction of Full and Apartment Railway Post Office Cars," dated July 20, 1938. Published by U. S. Post Office Department, Washington, D. C.

2 "Streamline Light-Weight High Speed Passenger Trains," by Coverdale and Colpitts, Report of June 30, 1938. Report available at 120 Wall St., New York, N. Y.

3 "Operation and Maintenance of Diesel and Steam Locomotives," by E. E. Chapman, presented before the Society of Automotive Engineers, October 13, 1938.

4 "Pioneering the Diesel Electric Streamliners," by Otto Jabelman, presented before Society of Automotive Engineers, Oct. 13, 1938.

5 "Light Weight Passenger Cars Used in Steam Railroad Service," by W. H. Mussey, presented before the Society of Automotive Engineers, October 13, 1938.

6 The following articles on lightweight trains published in *Railway Age*:

February 3, 1934, page 184, "Union Pacific Light Weight, High-Speed Passenger Train."

April 14, 1934, page 533, "Burlington *Zephyr*."

October 13, 1934, page 427, "Union Pacific Second High Speed, Light Weight Passenger Train."

December 22, 1934, page 825, "New York Central Streamlined *Commodore Vanderbilt*."

January 5, 1935, page 3, "New York, New Haven & Hartford Streamlined Coaches."

February 2, 1935, page 188, "North Western '400' High Speed Train."

February 9, 1935, page 220, "Maine Central, Boston & Maine *Flying Yankee*."

April 13, 1935, page 562, "Light Weight Coaches for Boston & Maine."

April 20, 1935, page 600, "Burlington *Twin City Zephyrs*."

April 27, 1935, page 632, "New York, New Haven & Hartford *Comet*."

May 4, 1935, page 671, "Baltimore & Ohio *Abraham Lincoln*."

May 11, 1935, page 719, "Chicago, Milwaukee, St Paul & Pacific *Hiawatha*."

June 8, 1935, page 875, "Union Pacific *City of Portland*."

June 15, 1935, page 910, "Gulf, Mobile & Northern *Rebels*."

November 2, 1935, page 563, "Chicago, Burlington & Quincy *Mark Twain Zephyr*."

March 28, 1936, page 534, "Illinois Central *Green Diamond*."

May 30, 1936, page 864, "Union Pacific *City of San Francisco* and *City of Los Angeles*."

July 4, 1936, page 4, "Union Pacific *City of Denver*."

July 11, 1936, page 50, "New York Central *The Mercury*."

December 26, 1936, page 931, "Enlarged *Twin City Zephyrs*, Chicago, Burlington and Quincy."

March 13, 1937, page 418, "Southern Pacific *Daylight*."

March 27, 1937, page 540, "New Haven Installs Streamline Passenger Locomotives."

May 22, 1937, page 855, "Santa Fe Re-Equips *Super-Chief*."

June 19, 1937, page 1013, "Richmond, Fredericksburg & Potomac 4-8-4."

August 28, 1937, page 256, "Chicago, Rock Island & Pacific Six *Rockets*."

October 2, 1937, page 442, "Southern Pacific *Sunbeam*."

December 11, 1937, page 826, "Reading Receives Light Weight Streamlined Train."

December 25, 1937, page 902, "Baltimore and Ohio Installs New *Royal Blue Train*."

January 15, 1938, page 147, "Gulf, Mobile & Northern's Third *Rebel*."

January 29, 1938, page 224, "*City of Los Angeles* and *City of San Francisco*."

February 26, 1938, page 372, "M. K. & T. New Passenger Cars."

March 26, 1938, page 554, "Santa Fe's New Streamlined Trains *El Capitan* and *Chief*."

May 14, 1938, page 838, "Chicago and North Western and Union Pacific *Challenger*."

June 4, 1938, page 942, "New York, Ontario & Western *Mountaineer*."

June 18, 1938, page 1000, "Pullman Builds New Equipment for *Broadway* and *Century* Trains."

October 15, 1938, page 559, "Union Pacific Streamliner Equipped With Additional Unit."

Discussion

T. R. Cook.[3] The writer agrees with the author's conclusions that the railroads have and can regain considerable of their lost passenger business by the adoption of better schedules and the provision of more comfortable passenger trains, and there is

[3] Coverdale and Colpitts, Consulting Engineers, New York, N. Y. Mem. A.S.M.E.

no question but that lighter construction is the principal fundamental factor involved.

As to whether these trains will be hauled by Diesel or steam power is a question which will be resolved by the economies of the situation. There are over 43,000 steam locomotives in existence in the United States, divided in the order of freight 65 per cent, passenger 17 per cent, and switchers 18 per cent.

It is quite evident that the major activity, that of hauling the freight will alone keep the steam locomotive on the rails for some years to come. A number of the railroads are coal carriers. It would, therefore, seem logical that the roads and the steam-locomotive builders, to say nothing of other investments in property and plants dependent upon steam, should constitute a sufficient urge to force the advancement of steam locomotives. There is no doubt but that there is room for improvement both in engineering and in operation of steam power.

As to the present situation of steam, the writer is doubtful of the superiority of the Diesel as indicated in the paper, particularly when the Diesel locomotive is considered as a prime mover. The author gives a thermal efficiency at the rail of 27 per cent for the Diesel and 7 per cent for the steam locomotive. The 27 per cent is correct for the Diesel, but 7 per cent in the writer's opinion is quite low for the modern steam locomotive.

A more interesting figure, however, would be a statement of the efficiency at the drawbar where the work is done, and not only at full load but at various load factors and at various speeds. We are dealing with types of power having entirely different characteristics. In order to make a comparative statement, the writer has worked out the thermal efficiencies of a modern 4-8-4 and a Diesel locomotive each of 5400 hp, through a range from 30 to 100 mph, and load factors varying from 100 per cent down to 40 per cent. The results indicate that the maximum efficiency of the Diesel is developed at 30 mph with 100 per cent load factor; this efficiency decreases as the speed increases and as the load factor decreases. The steam locomotive, on the other hand, generally increases in efficiency as the load drops and has higher efficiency at 80 per cent load at all speeds than at 100 per cent load; up to about 50 mph this increase in efficiency continues down to a 40 per cent load factor. It may fairly be assumed that, in high-speed passenger service, the load range is primarily between 60 and 80 per cent, at speeds varying from 50 to 70 mph. In the example taken in the ranges cited, the Diesel shows an efficiency of 22 per cent and the steam slightly over 8 per cent. This would indicate that the cost of the fuel in the normal range of unit prices would be about equal.

As to the cost of repairs, it is believed that the age of the equipment must be taken into consideration. The increased cost in repairs of steam locomotives with increasing age has been thoroughly demonstrated. The Diesel switcher has shown a greater rate of increase and, although there are no figures available, it is safe to assume that the Diesel road locomotive will show a similar increase.

The author in Fig. 1 shows comparative drawbar-horsepower curves of a steam versus a Diesel locomotive, stating that the curves are plotted from actual dynamometer-car readings. As a rule, dynamometer cars are placed in a given scheduled train and the readings taken show the resistance of the train under various conditions, but do not show the capacity of the prime mover at various speeds. The writer has found it of interest to compare the theoretical drawbar horsepower with that shown in the author's curves. His calculations indicated that the Diesel was operating below full load up to 80 mph and, thereafter, was working at an overload; while, in the case of steam, the locomotive was operating at full load up to 50 mph and was working considerably under capacity at higher speeds.

In Table 4 of the paper, in reference to steam power, what is

the author's definition of "maximum indicated horsepower?" The indicated horsepower of a steam locomotive increases with speed up to a point where the steam-pipe valve and port resistance prevents the use of all the steam the boiler can evaporate.

Under the section on "Maintenance," it is stated that the cost of repairs of a steam locomotive is 20 cents per mile, and the comparative Diesel costs 17 cents per mile; the writer would like to know the age of both the steam and the Diesel units used in this comparison.

Under the section on "Thermal Efficiency," a comparison is made between Diesel and steam power. In one case, a ratio of 5.5 is shown and in another a ratio of 4.5. In the two instances was the work done the same, i.e., did the Diesel and steam locomotives haul the same-weight trains over the same tracks?

L. B. Jones.[4] The author quite properly calls attention to the attractive weight reduction made possible by the use of lightweight cars. Parallel with the development of lightweight
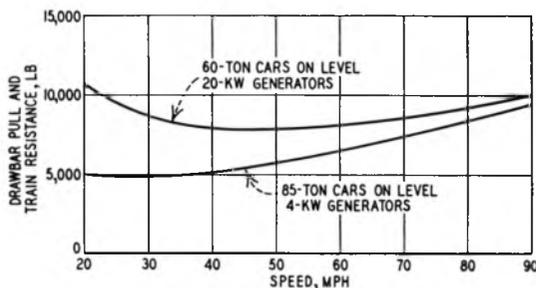

FIG. 4 DRAWBAR PULL REQUIRED FOR 12-CAR TRAINS
(85-ton cars, 1020 tons per train; 60-ton cars, 720 tons per train.)

cars, however, has come the development of mechanically driven air-cooling systems which offset any advantage the locomotive might realize from the substitution of lightweight cars in the train.

In Fig. 4, two 12-car trains have been assumed, one made up of 85-ton cars having 4-kw axle generators working to rated capacity, and another 12-car train of 60-ton cars, with 20-kw generators, generating at rated capacity. It will be noted that the drawbar pull required for the 60-ton cars is greater, up to speeds considerably in excess of 90 mph. The figures are for level track. On an upgrade, the lightweight cars would gain some advantage.

The development of heavy axle-generator loads has, therefore, robbed the locomotive of whatever advantages the lightweight cars offer in the way of drawbar pull. The development of air cooling has been very rapid, and sufficient consideration has not been given to a logical and comprehensive method of supplying the power necessary. As a starting point, it is suggested that a 440-v, 3-phase train line, powered from a small plant on the locomotive, would supply all necessary power for cooling and lighting at an over-all efficiency probably double that of individual axle-generator units. Storage batteries could also be reduced to about $1/4$ of their present size, and kept charged through a small automatic rectifier. Stand-by service would be provided in large terminals, as is now done with steam heat.

Reference is made to the lower center of gravity of Diesel-electric locomotives, with higher possible speed on curves. This is no doubt true as regards speeds at which the locomotive will upset, but there have been cases where it was necessary to raise the center of gravity of electric locomotives in order to prevent serious lateral forces which pushed the track out of line or over-

[4] Engineer of Tests, The Pennsylvania Railroad, Altoona, Pa. Mem. A.S.M.E.

turned a rail. In Figs. 5 and 6, curves have been plotted showing the maximum possible speed of stability of a locomotive on a 6-deg curve. Fig. 5 shows a curve with 5 in. superelevation and Fig. 6 a curve without superelevation.

While the speed required to upset the locomotive increases rapidly with reduced center of gravity, it will also be noted that the speed required to overturn the rail has the opposite trend, and the two curves cross at 98 mph in Fig. 5. These curves are plotted on the basis of centrifugal force only. When we consider the maximum forces resulting from nosing or lurching of the locomotive, it is evident that too low a center of gravity can create dangerous lateral pressure on the rail, and therefore the advantages of a low center of gravity must be circumscribed.
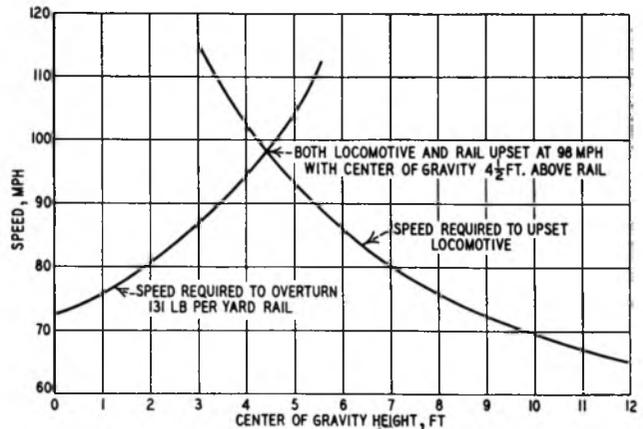
FIG. 5 CURVES SHOWING MAXIMUM POSSIBLE SPEED OF STABILITY OF A LOCOMOTIVE ON A 6-DEG CURVE WITH 5 IN. SUPERELEVATION
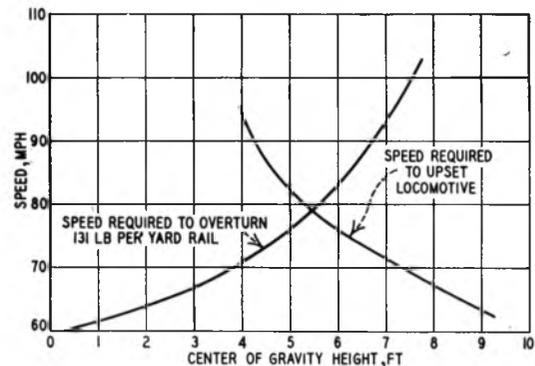
FIG. 6 CURVES SHOWING MAXIMUM POSSIBLE SPEED OF STABILITY OF A LOCOMOTIVE ON A 6-DEG CURVE WITHOUT SUPERELEVATION

Data are presented in the paper showing rail stresses under a steam locomotive in comparison with a Diesel. Omitting the steam-locomotive drivers for the moment, and considering only the engine and tender trucks, which are not subjected to dynamic forces, it will be noted that the rail stresses under the locomotive trucks increase much more rapidly from 5 to 100 mph than the stresses under the Diesel-locomotive trucks. It is not apparent why rail stresses should increase more rapidly with speed under a locomotive-tender truck, for example, than they do under a Diesel-locomotive truck.

Considering now the locomotive driving wheels, it will be noted that the minimum rail stress under the main wheel at 100 mph is only slightly less than the stress due to static loading at 5 mph (it is assumed that at 5 mph no dynamic stresses are noticeable). In most 2-cylinder steam locomotives which have been tested, there is a definite tendency for the main wheels to leave the rail

at a speed of 100 mph and, although with well-balanced loco-motives no lifting actually occurs, the rail stress at one point of the revolution should be very nearly zero. This raises the question whether the strain gages used were sufficiently sensi-tive to record the actual stress variations. If average figures were used in plotting the vertical stresses, this might explain the tendency to level off the variations throughout one revolution.

While there can be no challenging the author's statement that future development of the steam locomotive should look toward reduction of the dynamic augment, it seems only fair to observe that the comments on undesirable dynamic behavior have been based upon the performance of a few locomotives, whereas hundreds of other steam locomotives are handling fast schedules every day without any symptom of trouble. We recently con-ducted a series of tests in which several locomotives were slipped on greased track up to speeds very much in excess of their maxi-mum operating speeds and, while the dynamic effect was defi-nitely noticeable, there was no trace of rail damage in 100-lb rail on cinder ballast, and the rail stresses recorded were not excessive. The only precaution taken in connection with the locomotives tested was that the counterbalances were precisely checked in advance of the tests, as well as the quartering and tramming of the locomotive wheels.

A. I. Lipetz.[5] There is a common but unjustified notion that a Diesel-electric locomotive has the advantage of greater acceleration as compared with a steam locomotive. This is a remnant of the common theory that the electric locomotive provides quicker acceleration than the steam locomotive. In this consideration, the fact is overlooked that the electric loco-motive usually draws power from an outside source to about double its nominal power at low speeds. However, this is not true for the Diesel-electric locomotive, in which the electric feature lies only in the transmission, not in the generation of power. The advantage of the Diesel-electric locomotive is in the magnitude of the tractive effort only at low speeds, since the power is always limited by the power of the Diesel engines and by the greater number of drivers which are carried on the loco-motive chassis, just as the power of the steam locomotive is lim-ited by that of the boiler on the locomotive frame and its num-ber of drivers. Thus, the author's statement regarding the "rapid acceleration (of the Diesel-electric locomotive) because of the high tractive power at lower speeds" may be literally correct, but can be misleading if the speeds to which this applies are not given. This advantage of a Diesel locomotive is of compara-tively short duration and is not of the same order, as in the straight electric locomotive with an outside source of power. In order to evaluate this advantage, a more accurate and detailed calculation is necessary. This has been done and is presented in Fig. 7.

Referring to Table 4 of the paper, the steam locomotive shown in the third column is the one discussed by the author. The maximum indicated horsepower is given as 4300, and the graph, which is shown in Fig. 1, gives the drawbar horse-power, both for a steam and some Diesel locomotives. The steam locomotive is referred to as a modern locomotive of the 4-6-4 type. Regarding the drop of the horsepower shown in the curve, it is explained that this can be improved if a change both in the design of the valve and the valve gear is made. Fig. 7 also refers to the 4-6-4-type steam locomotive of the New York Central of a lighter design but, nevertheless, of 4700 ihp, built for high speed with this very improvement in the valves; the steam curve is taken from actual tests.

The drooping characteristic of the horsepower curve of the

[5] Chief Consulting Engineer, American Locomotive Company, Schenectady, N. Y. Fellow A.S.M.E.
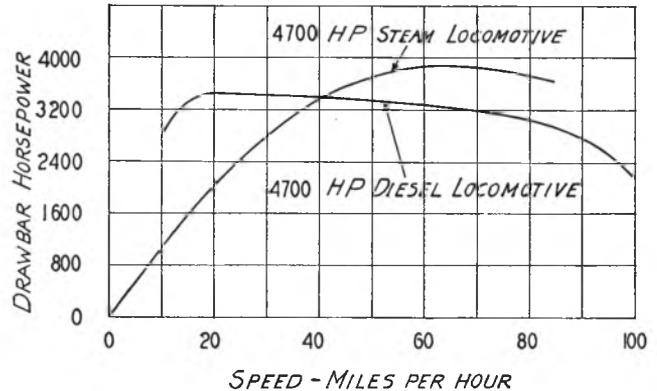
FIG. 7 COMPARISON OF DRAWBAR HORSEPOWER OF STEAM AND DIESEL LOCOMOTIVES OF IDENTICAL INDICATED HORSEPOWER
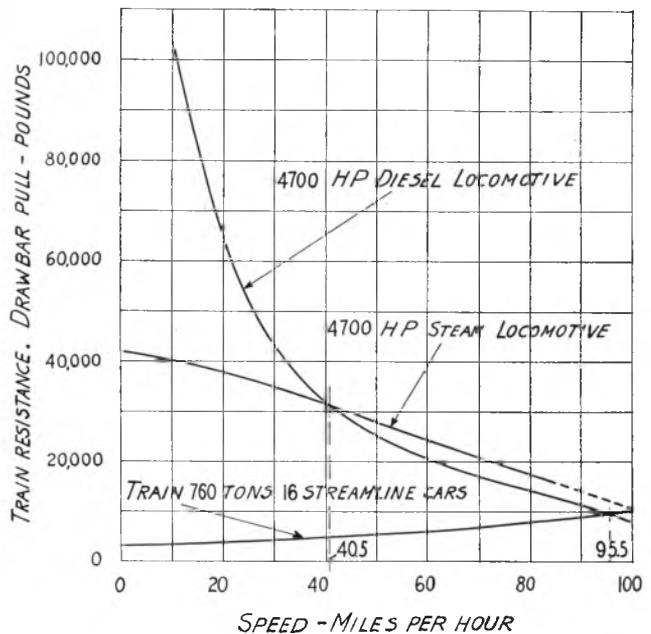
FIG. 8 TRACTIVE EFFORT IN RELATION TO SPEED IN MILES PER HOUR OF STEAM AND DIESEL LOCOMOTIVES OF IDENTICAL HORSE-POWER

steam locomotive of Fig. 1 has been corrected in Fig. 7; the part of the curve at high speeds has a flatter appearance. Further-more, the weight of the locomotive with a fully loaded tender is given in Table 4 of the paper as 809,000 lb. For the 4700-hp steam locomotive it is only 674,300 lb and, therefore, the weight per horsepower, instead of being 188 lb as shown in Table 4 is only 144 lb. Moreover it should be noted that the steam loco-motive is not running all the time with its tender fully loaded and, therefore, has the advantage of reducing its dead weight; this remains practically constant in a Diesel locomotive. How-ever, this slight advantage of the steam locomotive will be disre-garded and the resistance will be based upon the weight of the locomotive with fully loaded tender as given, namely, 674,300 lb. The maximum drawbar horsepower of the 4700-hp steam loco-motive is 3880 hp, as established by very elaborate tests[6] of the New York Central.

[6] "New New York Central Locomotives Show High Power Con-centration," *Railway Age,* vol. 104, 1938, p. 601. Also: "The New York Central Receives 50 Powerful 4-6-4 Locomotives," *Railway Mechanical Engineer,* vol. 112, 1938, p. 173.

As to the Diesel-electric locomotive shown in the second column of Table 4, the power of the Diesel engines is only 4000 hp and the maximum drawbar power is 2950 hp, against 3880 for the steam locomotive. In fairness to the Diesel-electric locomotive, which the writer will compare with the steam, the drawbar power given by the author in Fig. 1, from actual dynamometer tests of the 4000-hp Diesel locomotive, should be increased in proportion to the maximum indicated horsepowers of the steam and Diesel locomotives; in other words, for the Diesel-electric locomotive, the maximum drawbar power should be

$$2950 \times \frac{4700}{4000} = 3465 \text{ hp}$$

It is only fair to the Diesel-electric locomotive to assume that the Diesel power is equal to the indicated power of the steam locomotive, since both must be multiplied by the efficiency of power transmission and mechanical efficiency of the chassis, in order to arrive at the drawbar powers. Thus, we are impartially comparing a steam and a Diesel locomotive of identical maximum indicated horsepower, which is 4700 in either case, by studying their corresponding drawbar-horsepower curves.

Fig. 8 shows the tractive efforts in pounds in relation to speed in miles per hour for these two locomotives of identical maximum horsepower. The advantage of the tractive effort of the Diesel locomotive at low speeds is evident from this illustration. It is true that, due to the electric transmission there is an advantage in developing greater tractive effort than with the steam locomotive for all speeds below 40.5 mph, but the difference between these two curves drops very rapidly and is negative at speeds above 40.5 mph.

Also in Fig. 8 is shown the resistance of a 16-car lightweight streamlined train weighing 760 tons, or 47.5 tons per car. This is what the Diesel-electric locomotive would be able to pull on level track at high speed. The balancing speed of the Diesel-electric locomotive is about 95.5 mph on the level. For the steam locomotive it is slightly higher, about 101 mph, because of greater sustained power at high speed.

The acceleration curves of these two trains are shown in Fig. 9. From the comparison, it may be observed that the intersection point of 40.5 mph is reached by steam in 2 min 10 sec within a distance of 0.75 mile, whereas the Diesel will do it in 1 min 16 sec in a distance of 0.51 mile. The Diesel will thus gain 54 sec and 0.24 mile. But, we are concerned here with performances at higher speeds than 40.5 mph and, if we continue beyond 40.5 mph, we shall be gaining time with steam; in fact, the steam locomotive will reach 95.5 mph in 10 min 17 sec from the start, while the Diesel will take 15 min 20 sec to attain this speed. In other words, the Diesel will lose the 54 sec which it gained at the

start before reaching the intersection point of 40.5 mph, but will lose an additional 4 min 9 sec, showing a total loss from the start of 5 min 3 sec, just contrary to what might be expected. Every time the train is slowed down (due to speed restriction) to a speed about 40.5 mph and accelerated again, the total gain in time, in the case of the steam train, will be increasing cumulatively.

It is interesting to note that L. Dumas, assistant director of the French National Railway, who has had considerable experience with Dieselized streamlined trains, made a special report,[7] in which he gave a summary of three different reports before the International Railway Congress Association. In this report he also made a statement concerning the advantage of quick acceleration of trains due to Dieselization with electric transmission, as compared with steam locomotives, but confined his remarks to trains with electric-motor coaches, essentially rail cars, and high-speed trains of the German and French types, in which the ratio of Diesel power to the weight of the cars is very high. The acceleration is, therefore, due not to the electric transmission, but to the high power-to-weight ratio, which is lacking in the American high-speed trains with sleeping accommodations.

In connection with the author's remark that the performance of the steam locomotive, especially of modern type, depends largely on the engineer and is subject to increased efficiency and capacity, an interesting incident demonstrating this fact occurred on the Chicago, Milwaukee, St. Paul & Pacific on November 18, 1939. On one of the new class F-7 locomotives of the 4-6-4 type, which is now operating fast heavy trains out of Chicago, something went wrong and the locomotive was cut out at Milwaukee. A 4-4-2 *Hiawatha* locomotive, which is lighter and less powerful than the 4-6-4 regular locomotive, was put on in its stead and pulled the 10-car train, weighing approximately 1,000,000 lb from Milwaukee to Minneapolis, bringing it in on time and making up almost the entire delay of 22 min; the train arrived at St. Paul only 2 min late. It is interesting to note that originally the *Hiawatha* was designed to pull only 6 cars weighing 700,000 lb, but lately has pulled 8 cars of corresponding weight. A Diesel locomotive would meet its requirements only, and no more; it does not have the flexibility of the steam locomotive.

Referring to the matter of counterbalancing, discussed in the section on "Track Stresses," there can be no doubt but that the steam locomotive is at a disadvantage as compared with any other locomotive without dynamic augment, such as Diesel-elec-

---

[7] "Methods Used to Speed Up Passenger Trains, and Resulting Expenditure," Special Report by L. Dumas, Bulletin of the International Railway Congress Association (English edition), July, 1939, pp. 661–663.
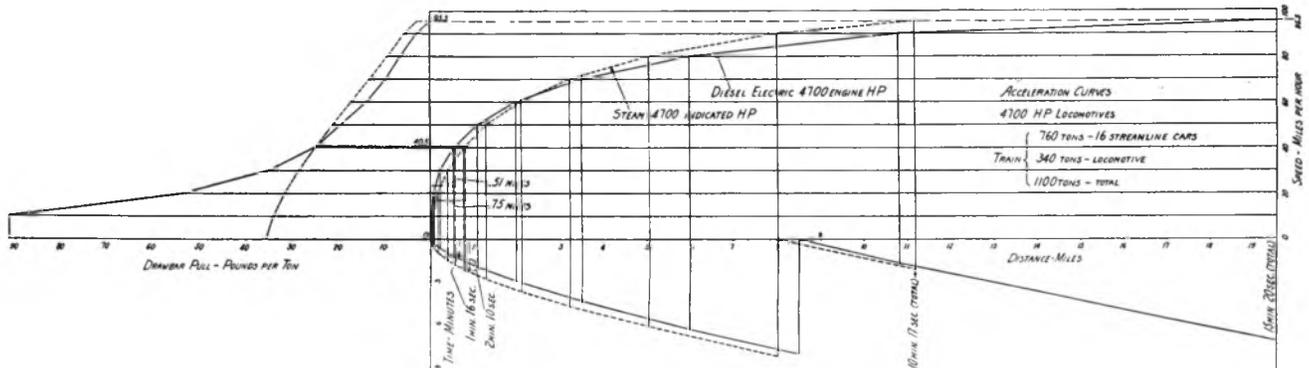

FIG. 9 ACCELERATION CURVES OF 4700-HP DIESEL-ELECTRIC AND STEAM LOCOMOTIVES
(Train composed of 16 streamlined cars, 760 tons; locomotive 340 tons; total 1100 tons.)

tric, turbine, or straight-electric. However, the examples given in the paper are not fully convincing; first, because the steam locomotive in question was heavy, 71,570 lb on the first driving axle, as compared with 50,700 lb on the fourth driving axle of the Diesel-electric. Furthermore, the locomotive, as may be judged from Fig. 3, was dynamically counterbalanced, or as we call it, cross-balanced, only on the main wheel, which was the second driving wheel. The front and back wheels were statically balanced. The percentage of balancing is not given and it is natural that the dynamic stresses in the rail should be greater for the steam locomotive. It is remarkable that, with all the handicaps of this steam locomotive, the difference is so small; the maximum rail stress is 19,000 lb for the steam locomotive at a speed of 100 mph as compared with 17,000 lb for the Diesel-electric at the same speed, or an increase of 12 per cent for the steam; much less than the increase in static weights, which is 41 per cent. For the Diesel-electric locomotive, the maximum at 100 mph is about the same as for the tender of the steam locomotive, although the weight on the tender axle is still about 30 per cent greater. This proves that the dynamic augment of the steam locomotive is not so grave as it is usually represented; it is the static weight which matters most, probably due to the element of time.

The old-time track and bridge engineers were mainly concerned about the static weight per axle. It is only recently, when the balancing of locomotives became better known and understood, that the dynamic stresses in rails began to bother them, especially when improperly evaluated.

The questions of track stresses and of counterbalancing of locomotives had a very peculiar history. Without going into details, it is enough to say that only lately proper attention is being paid to the question of counterbalancing of steam locomotives.

With the old-time counterbalancing, speeds up to 80 mph could be easily obtained. Hammer blows as high as 15,000 lb per wheel and more, which would kink and break rails, were evidently tolerated on a great many railroads. The track was simply built very stiff so the locomotives could run at these speeds. As one engineer expressed it, the "anvil" was strong enough to stand the "hammer blow." However, when the electric and the Diesel-electric locomotives came into being and proved their superiority for track conditions and their maintenance, the steam-locomotive engineer became cognizant of the unfitness of the old-time steam locomotive to modern high-speed conditions. Cross-balancing and lighter reciprocating weights, to which 15 years ago no one would listen, became very popular.

To a great extent the question of adaptability of steam locomotives to high speeds depends upon designing the locomotive so that it may have good tracking conditions at high speeds. American locomotives, especially of the type with front and rear trucks, have been very well suited to good negotiation of curves at high speeds. This partly eliminated the necessity of paying much attention to proper counterbalancing of locomotives. The recent development of lateral-motion and cushioning devices in locomotives was also a great help in this direction.

It is now possible to build locomotives with very small overbalances on wheels and, consequently, small hammer blows. Although it may not yet be possible to bring it down to zero, as was done on the Madras & Southern Mahratta Railway of British India, an improvement in track stresses, nevertheless, would result from the small overbalances, if the percentage of balancing is reduced to a figure close to zero and compatible with the shaking forces of light reciprocating weights in modern steam locomotives.

Wide discussion is taking place concerning the low center of gravity of some trains and Diesel-electric locomotives, as compared with the high center of gravity of steam locomotives, al-

though nothing is pointed out as to what kind of disadvantages may result from the high center of gravity of the steam locomotive. This question was investigated years ago by many locomotive designers. In the 1870's, it was even proved experimentally that the high center of gravity is an advantage because it stabilizes the locomotive at high speeds. Of course, on curves, the centrifugal force may overturn a locomotive, if proper care is not taken in the design, but if the correct track elevation is made and the centrifugal forces are thus neutralized, nothing will happen on a properly designed and maintained track. However, the locomotive will run more smoothly.

It should be remembered that any train, even on first-class track, is subjected to violent vibrations due to the unevenness of the permanent way, its discontinuity due to rail joints, the resiliency of rails, ties, and ballast, and that a certain limit of speed must be set for any train with any locomotive—steam, turbine, Diesel, or electric, on a flexible track.

In a recent paper,[8] F. G. Gurley, vice-president of the Atchison, Topeka & Santa Fe Railway, predicted speeds of 140 to 150 mph and indicated the motive-power equipment which would satisfy this condition. He even saw some possibilities for the steam locomotive, but, of course, had no doubts about the Diesel-electric locomotive at those speeds. As far as the writer is concerned, the figure of 150 mph cannot be accepted for any locomotive on any track until by actual experience the possibility of these high speeds is proved. It is not known that they are wanted for rail transportation.

It is interesting to note that electric locomotives on the Pennsylvania Railroad, which has first-class track, the heaviest in the world, are limited to 90 mph and are permitted to run at this speed only because their weight per axle is comparatively low.

Modern steam locomotives, properly cross-balanced on all axles, with cushioning devices on properly selected axles, can easily run at 100 mph, and have already proved their riding qualities at 115 mph. In fact, abroad, a steam locomotive developed 125 mph in the presence of the London Institution of Locomotive Engineers. References to this trip have been widely published.[9]

Speeds at 100 mph with steam locomotives are now not unusual in this country. There is no doubt in the writer's mind but that the steam locomotive can meet any higher speed, limited only by the speed of cars, obtained by any other type of locomotive.

In discussing the economics of the problem, the author in Table 6 gave us a comparison of operating costs between two streamlined trains in transcontinental service on one railroad. He also referred us to the well-known report prepared by Coverdale and Colpitts (2). He mentioned, however, that interest and depreciation charges have not been taken into account by these analysts.

This comparison has been completed by the writer by adding several more trains to Table 6, and by trying to figure out from various sources the cost of streamlined trains, using for this purpose a paper[10] by E. E. Chapman on the subject. The results of this comparison are given in Table 8. Several representative streamliners, like the *City of San Francisco*, *Denver Zephyr*, and *Hiawatha*, have been added and the table is complemented by the cost of locomotives and cars, fixed charges per year on the basis of 10 per cent of interest and depreciation, the income per train-

---

[8] "Performance Limits of Transportation; Range, Speed, and Capacity," by J. G. Gurley, Michigan-Life Conference, University of Michigan, Ann Arbor, Mich., November, 1939.

[9] "New German Streamlined Locomotives," *Railway Gazette*, vol. 62, 1935, pp. 1209–1217. Also: "The Borsig Centenary," *Railway Gazette*, vol. 67, 1937, p. 318.

[10] "Diesel and Steam Locomotives in High-Speed Service," by E. E. Chapman, *Railway Age*, vol. 105, 1938, pp. 733–735.

TABLE 8 OPERATING RESULTS OF HIGH-SPEED STREAMLINED TRAINS

	Southern Pacific	B. & O.		Union Pacific	Milwaukee	Burlington	Burlington Zephyrs (avg, 8 trains)	Santa Fe streamliners (avg, 7 trains)	Rock Island Rockets (avg, 6 trains)
Route	San Francisco–Los Angeles	Chicago–St. Louis		Chicago–San Francisco	Chicago–Twin Cities	Chicago–Twin Cities	....	....	....
Train	Daylight	A. Lincoln	A. Rutledge	City of San Francisco	Hiawatha	Twin City Zephyr			
Steam or Diesel-electric	Steam	Diesel-electric	Steam	Diesel-electric	Steam	Diesel-electric	Diesel-electric	Diesel-electric	Diesel-electric
Daily mileage per train	471	564	564	753	422	882	738	630	620
Train-miles per year	172000	204000	149000	275600	153000	322000	269000	229950	226300
Number of cars per train	14	8	8	14	9	7	5.90	6.85	3.33
Revenue (gross) per train-mile, dollars	4.631	3.020	2.585	4.188	4.343	1.902	1.893	2.072	1.343
Total operation cost per train-mile, dollars	1.420	1.038	1.065	1.659	1.126	0.619	0.610	0.801	0.501
Net revenue, dollars	3.211	1.982	1.520	2.529	3.217	1.283	1.283	1.271	0.842
Per cent of net revenue to gross	69.3	65.6	58.8	60.4	74.1	67.5	67.8	61.3	62.7
Cost of locomotive and train, dollars	1133000	711000	630000	1557000	713000	647000	569125	792292	379000
Fixed charges, 10 per cent total cost per yr, dollars	113300	71100	63000	155700	71300	64700	56913	79229	37900
Fixed charges per train-mile, dollars	0.658	0.349	0.422	0.565	0.466	0.201	0.211	0.345	0.187
Income per train-mile, dollars	2.553	1.633	1.098	1.964	2.751	1.082	1.072	0.926	0.655
Return on investment, per cent	38.8	46.9	26.0	34.7	59.0	53.8	50.7	26.9	39.1
Period of amortization in years	2.58	2.17	3.85	2.83	1.69	1.86	1.98	3.72	2.56

mile (net revenue less fixed charges), and amortization of the investment.

A study of Table 8 is illuminating. The number of cars is greatest only in the case of two trains, namely, in that of the *Daylight* on the Southern Pacific (14 cars) and the *City of San Francisco* on the Union Pacific (also 14 cars). The gross revenue of these two long trains amounts to more than $4 per train-mile, irrespective of whether Diesel or steam, thus indicating the obvious fact that the gross revenue depends upon the number of passengers, which naturally is greater on the longer trains. To the same class of large-revenue trains belongs the Milwaukee *Hiawatha* steam train, which has only 9 cars. This is due to its great popularity. All the Diesel trains, except the *City of San Francisco*, earn less gross income per mile.

The percentage of net revenue to gross is about the same in all streamliners (60 to 70 per cent), but the greatest is the *Hiawatha* (74.1 per cent), except the *Denver Zephyr* (included in the Burlington *Zephyrs*) for which it is 75.1 per cent. This is because the mileage of this train is very high, being more than double that of the *Hiawatha*. The return on investment, however, is lower.

The costs of the trains are, of course, different, the greatest being that of the long steam and Diesel-electric trains, especially the Diesel. This, of course, reduces the net revenues by the amount of fixed charges, making the income per train-mile lower for the Diesel trains. Only the steam *Daylight* and the steam *Hiawatha* show high incomes, above $2.50 per mile. The steam *Ann Rutledge* makes only about $1 per train-mile, probably due to traffic conditions. The total income per train-mile is figured, and the total per year is then referred to the investment. It will be seen that again the highest figure is for the *Hiawatha*, the cost of which is paid up in the shortest time, a little over 1 1/2 years.

The return on the investment of all these trains, both steam and Diesel-electrics, is not only the result of their utilization, especially the latter, but also a result of the comparatively low costs of the steam locomotives. In some cases the Diesel-electrics show remarkable mileage, as, for instance, the *Denver Zephyr*, which makes 1036 miles daily. However, but few trains, even in this country, have such a long run and can show such a remarkable utilization, which, after all, depends also upon the convenience of schedules, location of cities, meeting points, turnover of locomotives, and other factors. Therefore, the writer does not think the author's conclusion that Diesel trains will predominate is at all obvious, because to a great extent the success of Diesel trains is the result of novelty, streamlining, comfort, air conditioning, special service, etc., all of which are obtainable with steam trains and are matters of competition. In the long run

the preference is given on economic grounds, as pointed out by Mr. Gurley,[3] if the engineering of the product is right. There is nothing inherently wrong in the design of the steam locomotive, and this type of prime mover may be further perfected for high speed.

The writer agrees with the second part of the author's conclusion in which he states that radically new designs of steam locomotives will probably be built to meet high-speed operating conditions. He enumerates the possibilities. It should be added that the combustion-gas-turbine locomotive of Brown, Boveri & Company looms now on the horizon, and although it is not strictly a steam locomotive, it can be built with direct drive, which has been spelling the success of the steam locomotive for over a hundred years. Furthermore, it has no dynamic augment. So far, the steam locomotive has been able to meet new conditions in similar circumstances.

Of course, we cannot foresee the future. But there are no indications which would point away from the locomotive with direct drive.

K. F. NYSTROM.[11] In the cost of operation of lightweight trains, it is necessary to include the initial cost of the power plant and cars. In the writer's opinion the costs of steam locomotives and passenger-train cars, at the present time, are entirely too high, as the railroads, and particularly the builders, have failed to standardize on designs. It is appreciated that different operating conditions exist. However, both steam locomotives and passenger-train cars might be standardized to a point where the cost would be substantially reduced. If we are to continue making inroads on the private automobile and recapture passenger business for the American railroads, considerably greater economies will have to be made so that railroad fares can be placed at a popular-price level.

Much has been done in regard to reduction in weight of passenger-train cars, however, much remains to be done. Trucks under passenger cars delivered to railroads this year weigh 34,000 to 36,000 lb per carset. This weight, considering the total weight of the car, is too high. At the present time, one railroad is experimenting with a truck which will not weigh more than 25,000 lb per carset and appears to offer some decided advantages.

The author of the paper commented on the change in the Railway Post Office Department specification as follows: "This change will probably result in the addition of more material than is now used in center sills of lightweight cars designed with high center stills and buffers, with little compensating benefit from

---

[11] Chief Operating Officer, Chicago, Milwaukee, St. Paul & Pacific Railroad Company, Milwaukee, Wis. Mem. A.S.M.E.

added protection." The author, in his modesty, failed to mention that under his supervision a design[12] of passenger-train car was developed by the Federal Co-Ordinator of Transportation. This design was developed after exhaustive studies had been made and met all the requirements of the Railway Post Office Department's specification then existing. The estimated weight of this design was 96,900 lb and in the writer's opinion would have been entirely satisfactory.

Unfortunately, the Association of American Railroads prematurely adopted the so-called "tight-lock coupler" before a satisfactory application was developed, with the result that the car builders abandoned the conventional practice which had been in vogue for more than 20 years and lowered the center sills, taking all the stresses on the coupler instead of on the buffer and coupler. This change created an undesirable condition when new cars were coupled with older-type cars. To correct this condition, the specification for passenger cars was hastily changed and it is the writer's strong conviction that this specification is causing an undue penalty when employing material other than low-carbon or so-called high-tensile steel.

Three passenger cars were recently built for one railroad by three different carbuilders, one employing high-tensile steel, the second, stainless steel, and the third, aluminum. All three cars weighed substantially the same. The writer is satisfied that the reason for practically the same weight was not that one concern applied greater engineering skill than the other, but was solely due to arbitrary deflection requirements in the new specification. The foregoing is, in the writer's belief, substantial proof of the inconsistency in the existing specification and it is hoped that the Association of American Railroads will not stand in the way of progress, but correct this inconsistency.

J. W. RAGSDALE.[13] The author cites the case of a 79-ft coach which weighs 97,000 lb. To the uncalculating reader this weight may appear to be at variance with the weights more recently published (6). These figures approximate 104,000 lb. One is apt to overlook the additional statement that they also refer to coaches having a length of 84 ft 8 in. The difference in weight and the difference in length are almost exactly proportional.

The author also refers to the Association of American Railroads specification which emphasizes the strength of the center sill in car construction. He does not mention the novel requirement that deflection may govern instead of ultimate strength. While the writer is absolutely in sympathy with the spirit which inspired the new A.A.R. specification and with all of its other requirements, he has differed on this matter of deflection. It penalizes the use of high-tensile steels such as stainless steel and it penalizes the Budd type of construction. On the basis of strength alone, a stainless center sill of 8 sq in. would be sufficient. We are now using as much as 18 sq in. If we "floated" the center sill instead of making it an integral part of the structure, we could meet the strength requirement with 8 sq in. and have absolutely no deflection of the car body at all. But, this member should be made to serve a general structural purpose rather than to be a mere part of the draft gear. This belief costs us 2200 lb. This weight might be used to better advantage elsewhere.

After all, it is the resiliency of a structure which counts, not its deflection. An 800,000-lb end loading presumes a crash. A resilient structure might better withstand this crash than an absolutely rigid one. The A.A.R. limits the deflection under this loading to 1 in. or less. A truly resilient car structure might hump up in the middle as much as 4 in. and yet suffer no damage other than to the trim. When we have a wreck, trim is of small consideration compared to passenger safety. The writer, therefore, is opposed to the deflection clause of the A.A.R. specification.

Another phase discussed by the author, and again in the interest of safety, relates to braking. He indicates that any increased braking must automatically involve some means of sanding. While this is quite true, the writer wishes to point out that wheel locking results largely from the peaks of the braking-torque curve and that these peaks are far more pronounced with a metal shoe-to-wheel braking than they are with the disk type of braking, such as has been developed by the company with which he is connected. The eventual development may well be not only disk-type braking but automatic sanding and a torque control.

E. W. TEST.[14] The writer wishes to discuss several points, particularly the references involving stainless steel in comparison with the low-alloy high-tensile steels in car construction. The author states: "While it is true that the full use of the superior physical properties of stainless steel or the lightweight of aluminum is limited to some degree by deflection requirements, it should be possible, by using these materials instead of low-alloy steel, to construct a lighter car of equal strength."

The writer is prompted to discuss this statement, principally in view of the publicity which has been given stainless steel, in which the virtues of the metal have been widely heralded to the complete exclusion of its eccentricities. It is felt that, in an engineering discussion of this kind, utmost care should be exercised not to foster any questionable ideas which may have been implanted in the minds of readers by an overzealous publicity effort. Most readers do not have the opportunity carefully to study this enormous flow of publicity, hence, the value of such engineering papers as the present one.

The author's statement might be accepted without question by many, if consideration were given only to those physical properties of stainless steel which are superior. Tensile strength is the property which has been generally overemphasized, not only in the case of stainless steel, but of other metals as well.

A study of the tensile properties alone either of stainless steel, high-tensile low-alloy steel, or mild steel, as the basis for the selection of a material for passenger-car construction, is not the proper approach to such an important problem. A material with 4 times the tensile strength is not 4 times as strong. The word "strong" must embrace all the properties of a metal and the quality of each must be such that the metal will not fail under any of the conditions in which it is to serve.

A very important requirement of a metal, if it is to be used in a structure, is that it have high compressive-stability properties. Dr. C. S. Aitchison and Dr. L. B. Tuckerman of the National Bureau of Standards, in the introduction of their report No. 649 to the National Advisory Committee for Aeronautics, clearly expressed this idea as follows:

"During recent years, a remarkable expansion has taken place in the use of thin sheet and thin-wall material in lightweight structures, such as airplane wings and airplane fuselages. The strength of these structures is generally limited by the strength of certain members carrying compressive loads. These members have frequently been designed on the basis of the tensile properties of the material. This is convenient as the tensile test is relatively simple and is widely used. However, it may lead to an unsafe structure, on the one hand, or an uneconomical structure,

[12] "Design of Typical Lightweight Coach," Report of Mechanical Advisory Committee to the Federal Co-Ordinator of Transportation, Washington, D. C., 1935, pp. 597–678.

[13] Chief Engineer, Stainless Steel Division, The Edward G. Budd Manufacturing Company, Philadelphia, Pa.

[14] Assistant to President, Pullman-Standard Car Manufacturing Company, Chicago, Ill. Mem. A.S.M.E.

on the other hand, if the compressive properties of the material differ from the tensile properties. There is an urgent need for a method which makes possible a direct determination of compressive stress-strain graphs for thin-wall material."

Any material offered for lightweight car structures comes in this category and it is essential that the compressive stress-strain ratio of these materials be satisfactory so that sufficient stability may be assured. In case of a wreck, which must always be considered in the design of railway passenger cars, substantially every structural member may be called upon to resist high compressive stresses.

It is a well-known fact, at least for the last several years, that compressive tests on 110,000-lb per sq in. yield strength 18-8 stainless steel show a drop in the modulus of elasticity to 22,000,000 at stresses as low as 44,000 lb per sq in. At 57,000 lb per sq in. stress, the modulus of elasticity has dropped to 19,000,000. In other words, a very important factor with reference to the ability of the metal to stand up as a column has been reduced in the ratio of 29 to 19, as the unit compressive stress was increased from zero to 57,000 lb per sq in. Compressive tests further reveal that, long before the stress has reached the alleged yield point of 110,000 lb per sq in., the modulus of elasticity has dropped to 12,000,000. Incidentally, under increase of tensile stress, test samples also show a reduction of the modulus of elasticity. At a tensile stress of 88,000 lb per sq in. the modulus of elasticity has decreased to 22,000,000. This is the disappointing quality of stainless steel which is revealed by stress-strain graphs. It is a major deficiency and it must be a limiting factor in the design of a structure using this material.

We are not unmindful of the many good qualities of stainless steel but these, including its superior tensile strength, cannot be used to offset the deficiencies when the material is called upon to resist buckling, in which case the low and variable modulus of elasticity becomes the weak link in the otherwise strong chain.

A theoretical conclusion as to the weights of structures, based upon a comparison of the known properties of the metals involved, is at the best a precarious adventure, particularly so in the case of a passenger-car structure. It has been difficult, as the author stated, to make accurate weight comparisons of the different types of structures used in passenger cars because of the dissimilarity in specifications.

We have, however, one comparison available, which probably has been published since the paper was prepared. The following statement is quoted from the article[15] in question:

"Early in January the Pennsylvania ordered 15 dining cars, 5 from the Pullman-Standard Car Manufacturing Company, 5 from the Edward G. Budd Manufacturing Company, and 5 from the American Car & Foundry Company. All of the cars are built to the same general specifications and are believed to represent the first instance in which three dissimilar sets of materials and techniques of construction have been so employed."

In this instance, the stainless-steel car was approximately 5000 lb heavier than a similar car with a structure of low-alloy steel:

Stainless steel: 118,170 lb; designed and built by Edward G. Budd Manufacturing Company. Low-alloy steel: 113,240 lb; designed and built by American Car & Foundry Company.

It is felt that the author has dismissed too lightly the economic aspects of the stainless-steel car. If there have been instances where bids on cars with stainless-steel structures, including the variety of equipment, accessories, and trucks, have by some method been reduced to the total price of other cars having low-alloy-steel structures, it would appear that this total car price is

15 "Pennsylvania Receives Diners From Three Builders," *Railway Age*, vol. 107, 1939, p. 469.

of little significance in an engineering analysis of the relative costs of the two body structures. The cold facts are that stainless steel costs 30 to 40 cents more per lb than the low-alloy steels. If no decrease in weight can be obtained by the use of the high-priced metal, the body structure alone will be penalized for material only, as much as $8000 to $10,000 per car. Recent competitive bids have fully reflected this differential in cost of materials.

It has not been the writer's purpose to advance the opinion that stainless steel is not a suitable material for passenger-car structures. By the recognition of its deficiencies it can be successfully used. We doubt the possibility of producing a car of stainless-steel structure which will weigh less than a car of low-alloy high-tensile-steel structure and which will be of equal strength. Even though high-tensile stainless is nowhere used structurally in any important way in vehicles engaged in major transportation, except in the case of railway passenger cars, it is believed that no one disputes the propriety of such use. The one question in relation to its use is that of economics, namely, as to what it may prove to be able to do toward cheap maintenance in order to justify the substantially higher investment.

J. R. THOMPSON.[16] The writer has been with the Bureau of Valuation of the Interstate Commerce Commission since the valuation work started in 1914, and has also been assigned to cooperate with the Depreciation Section of the Bureau of Accounts since the depreciation order affecting equipment was put into effect in 1934. In this connection, they have been interested in the service life of all railway equipment in the United States and, in recent years, particularly in the new designs of locomotives and passenger-train cars. Therefore, it will be helpful if the author in his closure will comment further upon the life expectancy which modern Diesel locomotives and lightweight passenger-train cars may attain.

VICTOR R. WILLOUGHBY.[17] The A.A.R. passenger-car specification, wherein additional strength in the end portion of the center sills and the end sills is required, is in the writer's opinion a very necessary requirement, especially where the new lightweight equipment is operated in the same trains with the old heavier equipment having high center sills. It is true that this adds considerable weight to the underframe of the car but this additional weight is in the proper place and is of material value at the time of an accident. The author states that no cars exactly similar in type but of different materials and construction had been built.

Recently there has been built a lot of 15 dining cars;[15] 5 using aluminum, 5 using stainless steel and 5 using low-alloy high-tensile steel. These cars were all constructed to the same strength specifications, practically the same floor plans, and very similar equipment.

The interior design of the stainless-steel car added about 700 lb to its weight, as compared with the low-alloy high-tensile-steel car. The differences in specialties amounted to a saving on the stainless-steel car of about 200 lb, making a net weight credit for the stainless-steel car of 500 lb. The trucks were supposed to be identical. The dry weight of the stainless-steel car was nearly 5000 lb more than the low-alloy high-tensile-steel car, meaning that after adjustment, because of interior equipment and finish, the stainless-steel car weighed about 4500 lb more than the low-alloy high-tensile-steel car.

It is the writer's opinion that the modern passenger car, with

16 Equipment Engineer, Bureau of Valuation, Interstate Commerce Commission, Washington, D. C. Mem. A.S.M.E.

17 Vice-President in charge of Engineering, American Car & Foundry Company, New York, N. Y. Mem. A.S.M.E.

an over-all length of about 85 ft, meeting all the strength requirements of the A.A.R. specification, when equipped with 4-wheel trucks, will weigh in the neighborhood of 108,000 lb.

The most important problems, which yet require solution, are better riding trucks and improved braking.

W. E. WOODARD.[18] The author mentions the fact that steam-locomotive designers are studying the question of improving their designs, particularly for high-speed work, and out of this will probably come some valuable recommendations. He speaks of this as in the future. The writer will mention some results already accomplished in this direction.

Fortunately, the writer's experience relates to a locomotive which has a little less than 4000 ihp, so the data are comparable with the 4000-hp Diesel mentioned in the paper.

The method of comparing modern locomotive types, as used by the author, is open to question. For the steam locomotives the weights are given with a full load of coal and water. This is an unfair basis of comparison, so far as the steam locomotive is concerned, and makes the figures look extremely favorable for the Diesel. If we use two-thirds coal and water in the tender, the comparison can be made much more nearly on the basis of actual conditions of continuous operation. Moreover, this is the figure used in all tests[19] and comparisons of the Association of American Railroads.

Using this basis, the steam locomotive, to which reference has been made, compares as follows:

	Diesel locomotive	Steam locomotive 4-6-2 type
Nominal hp	4000	....
Maximum hp	.....	4000
Weight with two-thirds fuel and water, lb	609600[a]	506000
Weight of locomotive only, lb	.....	330800

[a] Latest published weights of 4000-hp Diesel.

Actual drawbar pulls have been secured from this 4000-hp steam locomotive and the curve looks very much different from that given in Fig. 1 of the paper. The steam-locomotive drawbar-pull curve crosses and goes above the Diesel at about 40 mph, touches almost 3000-drawbar hp at about 65 mph, and remains above the Diesel curve all the way to 100 mph. Such a drawbar-horsepower curve gives operating characteristics far better than the steam-locomotive curve in Fig. 1.

The author mentions the accelerating abilities of the Diesel locomotive as compared with steam locomotives. On all runs in passenger service, except strictly local, the ability to reaccelerate from 40 mph up to about 70 or 75 mph is far more important than the rapidity with which the train reaches 40 mph. Almost any division, say of 200 miles in length, has at least five or six slowdowns; it is the ability to reaccelerate at these points which counts. In other words, normally there are five or six slowdowns to one start from zero speed. This is the reason why the increased drawbar-horsepower curve of the 4000 ihp steam locomotive at 40 mph and above is of such vital importance, particularly in high-speed work.

As an example of what a 4000-hp steam locomotive is doing in regular service, some interesting runs recently made with a 13-car train, weighing approximately 900 tons, are mentioned as follows:

Distance traveled 98.3 miles at a speed of 83.07 mph
Distance traveled 78.8 miles at a speed of 84.4 mph
Distance traveled 25.1 miles at a speed of 100.4 mph

These figures are from the official train sheet. The maximum-

speed stretch of 25.1 miles was over rolling profile (not all downgrade).

In the case of a 12-car train, consisting of standard Pullmans and diners, weighing about 1100 tons, a maximum speed of 96 mph was attained on the run from Fort Wayne, Ind., to Crestline, O., which was accomplished in 2 hr flat, 5 min less than the "Broadway Limited's" schedule. One stretch of 116.4 miles was negotiated at a speed of 74.3 mph.

These are not mentioned as outstanding, simply because they were high-speed runs. Many such runs have been and are being made. The writer wishes to emphasize the fact that they were accomplished with a 4000-hp steam locomotive, the locomotive itself weighing 330,800 lb. This locomotive did not have any special crew or preparation. The runs were made in regular "pool service."

AUTHOR'S CLOSURE

Mr. Cook presents theoretical calculations indicating that the cost of fuel for Diesel and steam locomotives should be equal. The figures given in the paper are actual performance data for similar operations and the author feels that such data are more reliable than any estimates which might be made.

There is no basis for the theory that the maintenance expense for Diesel engines increases with the increasing age of the equipment. Thirty years' history with rail motorcars refutes this theory. Furthermore, the comparisons made by the author involve comparatively new steam locomotives as well as new Diesel locomotives. The steam locomotives referred to in the study of maintenance costs were built in 1928 and the figures were taken from records for the years 1930 to 1935. The figures used for the Diesel locomotives were taken from records for their first 4 years of service.

There is a misunderstanding of the methods mentioned as being used in making dynamometer-car tests. In testing a new type of locomotive, various train consists are used on numerous runs, so that the maximum performance can be checked. The steam locomotive referred to was not working below capacity at high speeds, but was being worked to maximum capacity. By maximum indicated horsepower the author refers to the maximum computed from the many hundreds of indicator cards taken during the various tests.

A question is raised regarding the comparison shown for thermal efficiency. The author made the comparison on exactly the same basis, that is, the same trains were handled on the same schedule.

Mr. Jones raises an interesting point as to the extra power required by the use of large axle generators, in connection with air-conditioning equipment. The author feels that the best answer to this problem is in the use of the steam-ejector system of air conditioning, wherein the power is secured from the locomotive boiler. The suggestion for the use of a head-end electric power plant to provide electricity for the cars may be suitable for certain conditions but its disadvantage lies in the fact that nonwired interchange cars could not be used in such trains.

It is undoubtedly true that locomotives with too low a center of gravity may produce serious lateral forces. It is noted from Mr. Jones's curves that the ideal height for the center of gravity is approximately 4 ft 6 in. This is the approximate height for Diesel locomotives, whereas the center of gravity for steam locomotives is considerably higher. The stress curves shown in Fig. 2 of the paper indicate that the Diesel locomotive does not produce any undue rail stress on curves. The question is asked why rail stress increases more rapidly with speed under locomotive tenders than is the case under Diesel locomotives. This is probably due to the fact that there is a greater shifting of wheel load in high-speed operation with steam locomotives than with

[18] Vice-President, Lima Locomotive Works, New York, N. Y. Mem. A.S.M.E.
[19] "A.A.R. Passenger Locomotive Tests," Report of Association of American Railroads, Chicago, Ill., October, 1938.

Diesel locomotives. Most track stress tests with steam locomotives have indicated this shifting of load.

The author does not agree with the inference that stresses developed under driving wheels are not serious unless rails are kinked. Higher stress must result in greater punishment to the rail and more chance of failure. It also results in increased road-bed-maintenance expense.

The curve presented by Mr. Lipetz for drawbar horsepower of a steam locomotive does not go beyond 85 mph and no figures are given to indicate how it droops beyond that point. His calculation of theoretical time for acceleration is interesting but in the dynamometer-car tests referred to in the paper different results were obtained. In mountainous territory, where there is sharp curvature, speeds are often brought below 40 mph. Furthermore, in service the steam locomotives do not appear to produce their theoretical power at all times. Most of the tests are made under the best operating conditions. The human element, the element of weather conditions, and the condition of locomotives all play an important part in steam-locomotive operation, whereas the performance of Diesel locomotives is much more uniform. There is no skill required to apply the power; rail conditions have very little effect because of nonslipping characteristics; and very cold weather is in no way detrimental to the Diesel. Another factor which appears to be a distinct advantage in the Diesel or electric locomotive is the continuity of the torque.

The author does not feel that the lack of cross-balance on other than the main wheels of a steam locomotive has any appreciable effect on track stress. It is not understood how Mr. Lipetz arrives at the figures quoted for maximum rail stress. It appears that there has been a misunderstanding of the diagrams Fig. 3, of the paper. The actual maximum stress at 100 mph is 27,000 lb for a steam locomotive and 17,000 lb for a Diesel locomotive at the same speed. Thus the increase for the steam locomotive is about 60 per cent instead of the 12 per cent referred to. It is undoubtedly true, as Mr. Lipetz states, that some locomotives have been built which may show a better performance than the one referred to in the paper. It is also true that a great many locomotives show poorer performance. It may be noted that the particular steam locomotive therein recorded was a strictly modern high-speed locomotive built in 1938, and therefore, should be fairly representative of present-day power.

As regards permissible speeds of operation on curves, the best indication of the facts is the definitely higher speeds permitted for Diesel locomotives than for steam locomotives on individual railroads. These rules have been based on actual experience.

As regards comparative costs of operation, the author does not believe it possible to compare operation on different railroads, because of the fact that operating conditions, traffic conditions, and equipment-maintenance conditions are different. It is for these reasons the author confined his comparisons to trains operating on one railroad.

The author does not agree that the success of the Diesel has been due to its novelty, steamlining, etc. In various parts of the paper the economic reasons for the success of the Diesel engine have been brought out. Economy of operation is the governing factor in the selection of any type of power.

The comments of Mr. Nystrom, Mr. Ragsdale, Mr. Willoughby, and Mr. Test indicate the wide difference of opinion which exists regarding materials and design best suited for lightweight passenger-car construction. The major parts of a passenger car must be designed to have strength characteristics far in excess of what is needed for ordinary operating conditions, in order to stand up in case of wreck. The best criterion is a study of wrecks which have actually occurred with equipment of various designs built with different types of lightweight materials and also the old-style heavy equipment. The author's observation of these results leads him to believe that the importance of deflection requirements has been exaggerated. What is desired in a passenger car is protection to lives of passengers in case of wreck rather than to minimize the damage to equipment. This feature should be the primary object in designing passenger cars. The author does not have the data upon which to base a comparison of the weights of the diners referred to in the discussion, but it is his understanding that there were certain requirements in the specifications which did not permit of the manufacturers designing the cars to make maximum usage of all the strength properties of the particular materials which they used.

Mr. Thompson requests further comments on service life of Diesel locomotives and lightweight passenger cars. There is no reason why this equipment should not have just as long physical life as the older types. However, there is a question as to how rapid will be the rate of obsolescence. Because of the Diesel power as well as the lightweight car construction being new developments, it is possible that there may be many improvements developed during the next 10 years, which would affect the rate of obsolescence to a greater extent than has been the case in the past with the older-type equipment, when there was much less new development work going on.