A Mineral Survey for Piezo-Electric Materials

By W. L. BOND

BECAUSE of the increasing interest in piezoelectric materials in many branches of science an exhaustive study of the minerals was undertaken with the object of finding all the materials that could possibly be of use for piezo-electric elements.   Much help was derived from existing data.[1]

Considerations of symmetry show us that for a crystal to be piezo-electrically active it must belong to a crystal class that has no center of symmetry (the Pentagonalicositetredral class of the cubic system, however, although it has no center of symmetry cannot be piezo active).[2]   This makes twenty classes of possible piezo activity and twelve classes that could not possibly be active.   About 90% of the crystals found in nature fall in those classes having centers of symmetry.

Although the mineralogical data are incomplete in their assignment of minerals to definite classes in the seven systems, the existing data give a start in the choosing of minerals likely to have useful piezo-electric properties.

All available data were gone through to obtain the following list of minerals classified by crystal structures.   As many of the non-centric ones as were obtainable in the United States were tested by the method of Geibe and Scheibe[3] (resonance in a thermionic oscillator circuit).   Whenever the authorities differed on the classification of a mineral it was so examined if obtainable.

In the mineral list, each mineral is numbered according to the number of the class in Groth's Physikalische Kristallographie, as follows: (*) indicating classes of possible activity:

*1 Assymmetric ⎫
 2 Pinacoidal ⎬ Triclinic system
             ⎭

---

[1] Dana—A System of Mineralogy, Ford—Dana's Textbook of Mineralogy; Groth—Chemische Kristalographie; Landölt Börnstein—Tabellen; International Critical Tables; Zeitschrift fur Kristalographie.

[2] W. Voigt, Kristal physik.

[3] *Zeits f Physik* 33, pg. 761 (1925).

*3 Sphenoidal
 4 Domatic   } Monoclinic system
 5 Prismatic

*6 Bisphenoidal
*7 Pyramidal   } Orthorhombic system
 8 Bipyramidal

*9 Bisphenoidal
*10 Pyramidal
*11 Scalenohedral
*12 Trapezohedral   } Tetragonal system
 13 Bipyramidal
*14 Ditetragonal Pyramidal
 15 Ditetragonal Bipyramidal

*16 Pyramidal
 17 Rhombohedral
*18 Trapezohedral
*19 Bipyramidal   } Rhombohedral system
*20 Ditrigonal pyramidal
 21 Ditrigonal Scalenohedral
*22 Ditrigonal Bipyramidal

*23 Pyramidal
*24 Trapezohedral
 25 Bipyramidal   } Hexagonal system
*26 Dihexagonal Pyramidal
 27 Dihexagonal Bipyramidal

*28 Tetrahedral-Pentagonal-Dodecahedral
 29 Pentagonal Icositetrahedral
 30 Dyakis-Dodecahedral   } Cubic system
*31 Hexakis-tetrahedral
 32 Hexakis Octahedral

In addition to the above classification, the following list of minerals is annotated with the following symbols:

A = active by test
I = inactive by test
R = unavailable or rare
M = mineral occurs only massive, amorphous or in other unsuitable form
S = crystal always very small
H = mineral is always non-homogeneous
U = unstable
C = electrically conducting
? = class not absolutely certain

CLASSIFIED LIST OF MINERALS

Actinolite	5	Allanite	5	Amosite	M
Adelite	5	Allemontite	21	Ampangabeite	8?UI
Aegirite	5?I	Allophane	M	Amphibole	5?HI
Aenigmatite	2	Almandite	32	Analcime	32
Aeschynite	8	Altaite	32	Ancylite	8
Alabandite	*31I	Aluminite	M	Andalusite	8
Alamosite	5	Alunite	21	Andesine	2
Albite	2	Alunogen	M	Andorite	8
Algodonite	H	Amblygonite	2	Andradite	32
Allactite	5?SI	Amesite	5	Anemousite	2

Mineral		Mineral		Mineral	
Anglesite	8	Bismite	21?I	Chillagite	10?
Anhydrite	8	Bismuthinite	8	Chloanthite	30
Ankerite	17	Bismutite	M	Chlorastrolite	H
Annabergite	5	Blödite	5	Chlorite	5
Annerodite	8	Blomstrandine	8?MI	Chloritoid	5
Anomite	5	Boleite	15?I	Chlormanganokalite	21
Anorthite	2	Boracite	*7A	Chloropal	M
Anorthoclase	2	Borax	5	Chloraphoenicite	I
Anthophyllite	8	Borickite	M	Chlorospinel	32
Antigorite	5?H	Bornite	*11I	Chondrodite	5
Antlerite	M	Boulangerite	8	Chromite	32
Apatite	25I	Bournonite	8	Chrysoberyll	8
Aphrosiderite	?I	Braunite	15	Chrysolite	8
Aphthitalite	21	Breithauptite	*20I	Cinnabar	*18I
Apophyllite	15	Britholite	27?S	Claudetite	5
Aragonite	8	Brochantite	8	Clausthalite	M
Ardennite	8	Bromyrite	32	Cleveite	32
Ardunite	M	Brookite	8	Clinochlor	5
Arfvedsonite	5	Brucite	21	Clinoclasite	5
Argentite	32	Brushite	5	Clinohedrite	*4A
Argentojarosite	I	Bunsenite	32	Clinohumite	5
Argyrodite	32	Bytownite	2	Clinozoisite	5
Arrhenite	H			Cobaltite	*28C
Arseniosiderite	8	Cabrerite	5	Cohenite	M
Arsonolite	32	Cacoxnite	M	Colemanite	5
Arsenophyrite	8	Calamine	*7A	Collinsite	I
Ascharite	M	Calaverite	5	Collophanite	M
Astrakanite	5	Calciothorite	M	Coloradoite	M
Astrophyllite	8?I	Calcite	21	Columbite	8
Atacamite	8	Caledonite	8	Connellite	25
Auerlite	15	Calomel	15	Cookeite	M
Augite	5I	Campylite	25	Cordylite	21
Aurichalcite	M	Cancrinite	27	Cornetite	?I
Automolite	32	Canfieldite	32	Corundum	21
Aventurine	2	Cannizzarite	?I	Corynite	28I
Axinite	2	Carnallite	8	Cotunnite	8
		Carnotite	I	Covellite	*18?I
Babingtonite	2	Carpholite	5	Crestmoreite	M
Baddeleyite	5	Caryocerite	21	Cristobalite	M
Baldaufite	?R	Cassiterite	15	Crocidolite	M
Barkevikite	5	Castorite	5	Crocoite	5
Barite	8	Caswellite	I	Cronstedtite	*16A
Barytocalcite	5	Catapleite	5	Crookesite	M
Bastnäsite	I	Celestite	8	Cryolite	5
Baumhauerite	5	Celsian	5	Cryolithionite	32
Bauxite	M	Cenosite	8?I	Cuprite	32
Beaverite	?S	Cerargyrite	32	Cuproscheelite	I
Bechilite	M	Cerite	8	Cyanite	2
Beckelite	32?S	Cerrusite	8	Cyrtolite	I
Bementite	8?I	Cervantite	8?		
Benitoite	*22I	Chabazite	21?I	Dahllite	M
Beraunite	I	Chalcanthite	2	Danburite	8
Bertrandite	*7I	Chalcedony	8?M	Datolite	5
Beryl	27	Chalcocite	8	Dawsonite	M
Beryllonite	8	Chalcolamprite	32	Dechenite	8
Berzelianite	MR	Chalcophyllite	21?I	Delessite	?SI
Berzelite	32?I	Chalcopyrite	*11C	Dellafosite	I
Betafite	32	Chalcosiderite	2	Delorenzite	8
Bindheimite	M	Chalcostibite	8	Delvauxite	M
Binnite	32?I	Chamosite	M	Demantoid	32
Biotite	5	Chiastolite	8	Deschloizite	8
Bischofite	5	Childrenite	8?I	Desmine	5

Deweylite	M	Freyalite	M	Heulandite	5
Diamond	31?I	Frieseite	8	Hielmite	8?I
Diaphorite	8	Fritzscheite	15	Hieratite	32
Diaspore	8	Fuchsite	I	Hillebrandite	M
Diopside	5			Hiortdahlite	2
Dioptase	17	Gadolinite	5	Hisingerite	M
Dixenite	?SI	Gageite	I	Hodgkinsonite	5?I
Dolemite	17	Gahnite	32	Hoeferite	M
Domeykite	8	Galena	32	Hokutolite	H
Douglasite	5	Ganomalite	I	Holmquistite	5?HI
Dufrenite	8	Garnet	32	Hopeite	8
Dufreneysite	5	Gastaldite	5	Howlite	M
Dumortierite	8	Gay-Lussite	5	Huebnerite	5
Dysanalyte	32	Gedrite	8	Humite	8
Dyscrasite	8	Gehlenite	15	Hussakite	*13
		Germantite	32	Hutchinsonite	8
Edingtonite	*6A	Gersdorffite	30	Hyalophane	5
Eleonorite	?R	Geyserite	M	Hydroboracite	5
Ellsworthite	MR	Gilsonite	M	Hydromagnesite	5
Elpidite	8	Gismondite	5	Hydrozincite	M
Embolite	32	Glaserite	21	Hypersthene	8
Emerald	27	Glauberite	5		
Emmonsite	?SI	Glaucodot	8	Ilmenite	17
Emplectite	8	Glauconite	M	Ilminerutile	15
Enargite	8	Glaucophane	5	Ilsemannite	M
Enstatite	8	Gmelinite	17	Ilvaite	8
Eosphorite	8?HI	Goethite	8	Inesite	2
Epidesmine	8?SI	Goslarite	*6I	Iodembolite	32?I
Epididymite	8	Graphite	21	Iodobromite	32
Epidote	5	Greenockite	*20IS	Iodyrite	*26?I
Epistilbite	*4?A	Griffithite	M	Iolite	8
Epistolite	5	Grossularite	32		
Epsomite	*6A	Guanajuatite	8?	Jadeite	5
Erikite	8	Gummite	M	Jamesonite	5?SI
Erythrite	5	Gymnite	M	Jarosite	21
Erythrosiderite	8	Gypsum	5	Jeffersonite	5?I
Euclase	5			Jenkinsite	M
Euchroite	8?I	Hackmanite	I	Jezekite	5?
Eucolite	21	Haidingerite	?S	Johnstrupite	5
Eucairite	M	Halite	32	Jordanite	5
Eudialyte	21	Halloysite	M	Joseite	M
Eudidylite	5	Hambergite	8		
Eulytite	*31I	Hancockite	5?S	Kainite	5
Euxenite	8	Hanksite	27	Kalinite	30
		Hardystonite	M	Kaolinite	5
Fairfieldite	2	Harmotone	5	Kasolite	I
Fassaite	5	Hatchettolite	32?I	Kelihauite	5
Faujasite	32	Hauerite	*28I	Kentrolite	8
Fayalite	8	Hausmannite	*11I	Kermesite	5?SI
Ferberite	5	Hauynite	*31I	Kieserite	5
Fergusonite	*10I	Hedenbergite	5	Klaprotholite	8
Ferrierite	I	Hedyphane	M	Klebelsbergite	?S
Florencite	21	Heintzite	5	Knopite	32?I
Fluocerite	27	Hellandite	5	Kobaltmanganerz	M
Fluorite	32	Heloite	*28?R	Koenenite	21
Forsterite	8	Helvite	*31I	Koppite	32
Forshagite	M	Hematite	21	Kornerupine	8
Fouquerite	I	Hercynite	32	Krennerite	8
Fowlerite	2	Herderite	8	Kroehnkite	5
Francolite	25	Herrengrundite	5	Kunzite	2
Franklinite	32	Hessite	32		
Freibergite	*31C	Hetaerolite	M		

Labradorite	2	Melanite	32	Noselite	*31I
Langbanite	17	Melanocerite	21	Nowmeite	M
Langbeinite	*28A	Melanophlogite	?SI		
Langite	8	Melanterite	5	Ochrolite	I
Lanthanite	8	Melilite	15	Octahedrite	15
Lapis-lazuli	H	Meliphanite	*9?A	Okenite	M
Laumontite	5	Mellite	15	Oligoclase	2
Laurionite	8	Mendozite	30	Olivenite	8
Laurite	*28	Menilite	M	Olivine	8
Lautarite	5	Merwinite	I	Omphacite	M
Lavenite	5	Mesolite	5	Onofrite	31
Lawsonite	8	Metacinnabarite	*31I	Opal	M
Lazulite	5	Meta Torbernite	I	Orpiment	8
Lazurite	5	Metavoltine	?SI	Orthoclase	2
Leadhillite	5	Miargyrite	5	Osmiridium	21
Lehnerite	I	Microcline	2	Otavite	21
Lehrbachite	M	Microlite	32	Ottrelite	2?I
Leonite	5	Microperthite	?HS		
Lepidolite	5	Microsommite	?SI	Pachnolite	5
Lepidomelane	H	Miersite	*31R	Pandermite	5
Leucite	*31?I	Milarite	27*	Paragonite	5
Leucophanite	*6A	Millerite	*20I	Parahoepite	2
Leucophoenicite	5?I	Mimetene	25	Paralaurionite	5
Libethenite	8	Mimetite	25, 231	Paratakamite	21?
Limonite	M	Minium	?S	Paravavxite	I
Linarite	5	Mirabilite	5	Pargasite	5
Linnaeite	32?	Mizzonite	13	Parisite	21
Licroconite	5?I	Molybdenite	27	Patronite	M
Liskeardite	M	Molybdite	8	Pearceite	5
Lithiophilite	8	Monazite	5	Pectolite	5
Loeweite	15	Monticellite	8	Penninite	5
Loellingite	8	Montmorillonite	M	Pentlandite	32
Loparite	I	Montroydit	8	Percylite	32?I
Lorandite	5	Morensonite	6	Periclase	32
Loranskite	8?	Morganite	27	Peristerite	2
Ludlamite	5?I	Mosandrite	5	Perovskite	8?
Ludwigite	M	Mossite	15	Perthite	?H, S
		Mottramite	M	Petalite	5
Magnesite	21	Muellerite	M	Petzite	32?
Magnetite	32	Muscovite	5	Pharmacolite	5
Magnetoplumbite	I	Muthmannite	*7R	Pharmacosiderite	*31I
Malachite	5			Phenacite	17
Malacon	I	Nadorite	8?I	Phillipsite	5
Mallardite	MI	Nagyagite	8	Phlogopite	5
Manganhedenbergite	5?I	Natrolite	8	Phosgenite	15
Manganite	8	Natron	5	Phosphoferrite	M
Manganophyllite	I	Naumannite	32	Phosphophyllite	5
Manganosite	32	Nemalite	M	Phosphosiderite	I
Manganotantalite	8?	Neotantalite	32	Phosphuranylite	M
Marcasite	8	Neotocite	M	Pickeringite	M
Margarite	5?RI	Nephelite	*23I	Picotite	32
Margarosanite	2	Nephrite	M	Picromerite	5
Margasite	5	Neptunite	5	Piedmontite	5
Marialite	13	Nesquehonite	8?I	Pinakiolite	I
Marignacite	32	Niccolite	*20I	Pinguite	M
Marmolite	M	Nickolsonite	8	Pinite	M
Marshite	*31I	Nickelbluete	5	Pinnoite	*10I
Martite	32?I	Nickeleisen	32	Pirrsonite	*7
Mascagnite	8	Niter	8	Pisolite	M
Matlockite	15?I	Nocerite	21?SI	Pitchblende	32
Maucherite	15?I	Northrupite	32	Plagionite	5?I
Meionite	15I			Plattnerite	15

Pleonast	32	Romeite	32?I	Stilpnosiderite	M
Plumbojarosite	21	Roscoelite	?S	Stolzite	13
Polianite	15	Rosenbushite	5	Strengite	8
Pollucite	?I	Rowlandite	M	Stromeyerite	8
Polybasite	5	Ruby	21	Strontianite	8
Polycrase	8	Rumpfite	M	Struvite	*7A
Polydymite	32?	Rutherfordine	5	Sulfoborite	8?R
Polyhalite	5?I	Rutile	15	Sulfur	8 or 7
Polymignite	8			Sulvanite	M
Powellite	13C	Safflorite	8	Sussexite	M
Prehnite	*7I	Sal-ammoniac	*28	Svanbergite	21
Priorite	8?I	Salite	5	Sychnodymite	32
Prismatine	8	Samarskite	8	Sylvanite	5
Probertite	M	Sanidine	5	Sylvite	*28I
Prochlorite	5?I	Sapphirine	5	Symplesite	?SI
Proustite	*20C	Sarcolite	*10?I	Syndalphite	5?
Pseudobrookite	8?S	Sartorite	5?S	Synganite	5
Pseudomalichite	M	Sassolite	2		
Psilomelane	M	Scheelite	13I	Tachyaphaltite	15
Psittacinite	M	Schefferite	5	Tachyhydrite	21M
Ptilolite	?S	Schirmerite	M	Talc	5
Pucherite	8	Schizolite	2	Tantalite	8
Pumpellyite	I	Schorlomite	32	Tapiolite	15
Pyroargyrite	*20I	Schreibersite	M	Tarbuttite	2
Pyrite	30	Schrockingerite	8	Tasmanite	?S
Pyroaurite	21	Schrotterite	M	Teallite	8?S, I
Pyrochlore	32	Schwartzengergite	?S	Tengerite	M
Pyrochroite	*20?I	Schwetzite	*31	Tennantite	*31I
Pyrolusite	8?HI	Scolecite	*4A	Tenorite	8
Pyromorphite	25	Scorodite	8	Tephroite	8
Pyrope	32?	Semseyite	5?I	Tetradymite	21
Phyrophanite	17	Senarmontite	32	Tetrahedrite	*31C
Pyrophyllite	8	Sepiolite	M	Thalenite	5
Pyropissite	M	Serpentine	5	Thaumasite	M
Pyrosmallite	I	Serpierite	8?S	Thenardite	8
Pyrostilpnite	5	Shortite	*7A	Thermonatrite	8
Pyroxene	5	Siderite	21	Thomsenolite	5
Pyroxmangite	2	Sillimanite	8	Thomsonite	8
Pyrrhotite	*20?C	Sipylite	*10I	Thorianite	32
		Skemmatite	H	Thorite	15
Quartz	*18A	Skutterudite	30	Thortveitite	8?I
Quenselite	I	Smaltite	*38I	Thuringite	M
Quercyite	M	Smithsonite	21	Tiemannite	*31A
Quisqweite	M	Sodalite	*31?I	Tiger-eye	M
		Sodaniter	21	Tilasite	5?I
Ralstonite	32	Spencerite	5	Titanite	5?I
Rammelsbergite	8	Spessartite	32	Titanmagneteisen	32
Raspite	5	Sphalerite	*31A	Topaz	8
Realgar	5	Spinel	32	Topazolite	32
Rhabdophanite	M	Spodumene	5	Torbernite	15
Rhodochrosite	21	Spurrite	5?I	Tourmaline	*20A
Rhodolite	I	Staffelite	M	Trechmannite	17
Rhodonite	2	Stannite	*11I	Tremolite	5
Rhomite	2	Staurolite	8	Tridymite	32?I
Richterite	I	Steenstrupine	21	Trimerite	2
Richardite	M	Stephanite	*7RI	Triphylite	5
Riebeckite	5	Sternbergite	8	Triplite	5
Rinkite	5	Stibiconite	M	Triploidite	5
Rinneit	21	Stibiotantalite	*7A	Tritomite	21?
Ripidolite	5	Stibnite	8?I	Troegerite	5?
Risorite	32?I	Stichtite	I	Troilite	M
Riversideite	M	Stilbite	5?	Trona	5

Troostite	17	Variscite	8?S	Wollastonite	5
Tscheffkinite	M, H	Vauxite	I	Wulfenite	*10
Tschermigite	30	Vermiculite	I	Wurtzite	*20A
Tungstenite	M	Vesuvianite	15		
Tungstite	8	Villiaumite	32?I		
Turgite	I	Vivianite	5	Xanthoconite	21
Turquois	2	Volborthite	I	Xanthophyllite	5?I
Tychite	32	Voltaite	32?I	Xanthoxenite	5?S
Tyrolite	M	Vonsenite	8?I	Xenotime	15
Tysonite	27				
		Wad	M	Yttrialite	M
Ulexite	M	Wagnerite	5	Yttrocerite	M
Ullmannite	30	Walpurgite	2	Yttrofluorite	32
Uralite	?HI	Warwickite	I	Yttrokrasite	8
Uraninite	32	Wavellite	I	Yttrotantalite	8
Uranocricite	8?	Wernerite	*10I		
Uranophone	M	Whewellite	5		
Uranopilite	M	Whitneyite	MI	Zeratite	M
Uranosphaerite	M	Wiikite	I	Zaophyllite	I
Uranospinite	8?	Wilkeite	I	Zeunerite	15
Uranothallite	8	Willemite	17	Zincite	*19?I
Uranothorite	M	Wilsonite	I	Zinkenite	8
Uranotile	2	Witherite	8	Zinwaldite	5
Utahite	?S	Wittichenite	8	Zircon	15
Uvanite	8?S	Woehlerite	5	Zirkelite	32
Uvarogite	32	Wolfachite	8	Zoisite	8
		Wolframite	5	Zorgite	M
Valentinite	8			Zunyite	*31AS
Vanadinite	25I				

Of the 830 minerals listed 70 belong to classes that allow piezo-activity but only 17 are found to be active by the Giebe and Scheibe test. (Our test of Iodyrite was negative but Greenwood and Tomboulian[4] found it to be active; on the other hand, we found Scolecite to be active while they report it inactive.) It may be that others of the remaining 56 classes have such small piezo-electric constants as to be undetectable. Others may be incorrectly classified as to symmetry.

Of these active materials, quartz is the most important. Because of its excellent mechanical properties (stability, etc.) as well as for its relative cheapness it seems destined to remain one of the most important piezo materials.

Tourmaline is also important because of the high magnitude of its elastic moduli in certain directions; however, it cannot be obtained in large pieces of satisfactory homogeneity.

Sphalerite is very difficult to handle because of its many cleavage planes, and appears to give little promise of becoming practically useful. Its activity is quite marked.

Homogeneous crystals of calamine appear to be very rare, so that workable crystals large enough for ordinary piezo-electric application are unobtainable. Most of the material occurs massive.

[4] On Piezo Electricity—Greenwood and Tomboulian—Zeits. f. Krist. Jan. 1932.

Epsomite gives a marked response but the crystals are generally small and they do not weather well.   There is some possibility, however, that they can be made artificially.

Boracite gives a marked response, but boracite alters slowly.   Its impermanence may bar it for some uses.

Stibiotantalite occurs only in thin scales, and the necessary cuts must be made in the most wasteful way.   Twinning is prevalent and the composition varies widely.

Scolecite occurs only as small crystals a few millimeters in diameter and a centimeter or so in length, uniformly twinned.

Iodyrite has been found to be active by other investigators.   It is electrically conductive, very soft and not very common.

Struvite is soft, unstable, and occurs only in small crystals.

Zunyite occurs only in minute crystals.

Langbeinite slowly changes its crystal structure.   It may be made artificially so may be of some use if it can be kept from alteration.

Leucophanite and Meliphanite are related minerals.   Neither seems to occur in good (i.e., homogeneous and untwinned) crystals of usable size.

Wurtzite does not appear very active but *good* crystals were not obtainable.

Tiemannite crystals were also unobtainable, but fragments of massive tiemannite responded.   Crystals might respond more energetically if they were obtainable, but minerals that are too difficult to get would not be of practical use.

Epistilbite occurs only in small specimens, uniformly twinned.

The mineral clinohedrite is strongly active but crystals are very rare.

Cronstedtite and Edingtonite are very weakly active.   Crystals of these are very rare.

The Fundamental Equations of Electron Motion (Dynamics of High Speed Particles)

By L. A. MacColl

I. Introduction

In work relating to the motion of electrons and other particles it is fairly common to assume that the particles obey the laws of Newtonian dynamics. That is, briefly, it is assumed that the rectangular coordinates (x, y, z) of the particle under consideration satisfy the differential equations

$$m\ddot{x} = X, \qquad m\ddot{y} = Y, \qquad m\ddot{z} = Z,$$

where m is the mass of the particle (assumed constant), X, Y, and Z are the components of the applied force, and the dots indicate differentiation with respect to the time t.

However, it is well recognized now that the above equations are not strictly correct, and that they merely represent an approximation which is adequate when the speed of the particle is sufficiently small compared with the speed of light. The system of dynamics based upon the correct equations[1] (which will be exhibited presently) is commonly called *relativistic dynamics*, not because any knowledge of the theory of relativity is essential to its understanding and use[2], but because it is in agreement with the theory of relativity (which Newtonian dynamics is not), because it was first developed in connection with work on the theory of relativity, and because even yet virtually all of the expositions of the subject are to be found in books and papers dealing primarily with the theory of relativity.

Just where the dividing line should be set between cases in which Newtonian dynamics is an adequate approximation and cases in which it is necessary to use relativistic dynamics is, of course, a rather vague question which cannot be answered simply and definitely. We may note, however,

---

[1] It is not the purpose of this article to discuss questions of fundamental physics, or the physical validity of any particular equations. For purposes of discussion, we assume outright that relativistic dynamics is at least more nearly correct than is Newtonian dynamics.

[2] The theory of relativity can be described briefly as a theory of the relations between the descriptions of phenomena in terms of different systems of reference. We shall *not* be concerned with this theory, because we shall be employing the same reference system throughout most of our discussion. In the final section of the paper we shall consider purely geometrical transformations of the coordinate system. These transformations, however, involve nothing that is really characteristic of the theory of relativity in the usual sense.

that according to relativistic dynamics the mass of a five thousand volt electron is about one per cent greater than the mass of an electron at rest. From this we can infer that, while Newtonian dynamics may be adequate for many purposes in our studies of electron motion, we do not have any great amount of margin, and that it will be necessary to use relativistic dynamics whenever we wish to obtain really good results concerning the motion of even moderately high speed electrons.

This article is purely expository. Its purpose is to set forth the fundamental equations and theorems of relativistic particle dynamics in a clear and concise form, unencumbered with any material relating to the theory of relativity proper. Almost all of the material is to be regarded as already known, but apparently it is only to be found in an inconvenient and scattered form. The incomplete bibliography at the end of the paper gives references to some of the more accessible sources of this and other related material.

II. The Elementary Differential Equations of Motion

Our discussion might be begun in any one of a number of ways, and no doubt the different approaches would appeal unequally to different readers. Considering the nature and purposes of this article, the author has deemed it best to begin by writing down at once the differential equations of motion of a particle (according to relativistic dynamics) in their most elementary form. Then, for the purposes of this discussion, these equations will have the status of a fundamental assumption. It need hardly be said that the equations are not written down arbitrarily. On the contrary, they represent the consensus of modern opinion as to the laws under which particles really do move.[3] The grounds, experimental and theoretical, for this opinion are set forth in various of the works cited in the bibliography.

For the time being, until the contrary is stated in the final section, we employ a fixed rectangular coordinate system. Instead of denoting the coordinates of the particle by x, y, and z, as we have done provisionally in the Introduction, we shall denote them by x_1, x_2, and x_3. Then $\dot{x}_1$, $\dot{x}_2$, and $\dot{x}_3$ denote the components of the velocity of the particle. The components of the force acting on the particle will be denoted by X_1, X_2, and X_3. For the time being we need only note that the force may depend upon the coordinates, the velocity, and the time; later on we shall introduce some more explicit assumptions about the force. The symbol c will be used to denote the speed of light in vacuo.

---

[3] The validity of these laws is not unrestricted. It is limited on the one hand by the quantum phenomena which become appreciable on the atomic scale, and on the other hand by certain phenomena revealed by the general theory of relativity which become appreciable on the cosmic scale.

We assume that the particle moves, under the influence of the force (X_1, X_2, X_3), so that its coordinates satisfy the system of differential equations

$$\frac{d}{dt} \frac{m_0 \dot{x}_n}{\sqrt{1 - (v^2/c^2)}} = X_n, \qquad (n = 1, 2, 3), \tag{1}$$

where m_0 is a positive constant characteristic of the particle, and v^2 is an abbreviation for the expression $\dot{x}_1{}^2 + \dot{x}_2{}^2 + \dot{x}_3{}^2$.* The positive value of the square root is the significant one; and wherever square roots appear in the subsequent work it will be understood, unless the contrary is stated, that the positive values are intended.

A few remarks may help bring out the significance of the foregoing assumption and its relations to the corresponding fundamental assumption of Newtonian dynamics.

We call the constant m_0 the *rest-mass* of the particle, and we assume (in accordance with the experimental evidence) that m_0 is identical with the mass of the particle which is used in Newtonian dynamics. In relativistic dynamics the quantity m defined by the equation

$$m = \frac{m_0}{\sqrt{1 - (v^2/c^2)}}$$

is called the *mass* of the particle. We note that as v/c approaches zero the mass approaches the rest-mass (whence the appropriateness of the latter term), and that as v/c approaches unity the mass increases without limit.

Consider the vector having the components p_1, p_2, p_3 defined by the formulae

$$p_n = \frac{m_0 \dot{x}_n}{\sqrt{1 - (v^2/c^2)}}. \tag{2}$$

We call this vector the *momentum* of the particle. The momentum is equal to the velocity of the particle multiplied by the mass.

Now equations (1) assert that the time-rate of change of the momentum of the particle is equal to the applied force.

We have already observed that as v/c approaches zero the relativistic mass of a particle approaches the Newtonian mass. We now note that as v/c approaches zero the components of the relativistic momentum approach the values

$$p_n = m_0 \dot{x}_n, \tag{2'}$$

* We might merely say that v is the speed of the particle. However, for our immediate purposes, it is important not to lose sight of the fact that v is a certain particular function of the components of velocity.

which are precisely the components of the momentum according to the Newtonian theory.

Finally, as v/c approaches zero, the differential equations of motion (1) approach the forms[4]

$$\frac{d}{dt}(m_0 \dot{x}_n) = X_n, \tag{1'}$$

which are the Newtonian differential equations of motion.

Thus we see that Newtonian dynamics is in effect a simplified approximate form of relativistic dynamics which is valid when the speed of the particle under consideration is sufficiently small compared with the speed of light.

Let us carry out the indicated differentiations in equations (1), and then solve the resulting equations for the quantities $m_0\ddot{x}_1$, $m_0\ddot{x}_2$, $m_0\ddot{x}_3$. The work is straightforward, and need not be given here. We obtain the following set of formulae:

$$m_0 \ddot{x}_1 = (1 - v^2 c^{-2})^{-1/2} \begin{vmatrix} X_1 & \dot{x}_1 \dot{x}_2 c^{-2} & \dot{x}_1 \dot{x}_3 c^{-2} \\ X_2 & 1 - (\dot{x}_1^2 + \dot{x}_3^2)c^{-2} & \dot{x}_2 \dot{x}_3 c^{-2} \\ X_3 & \dot{x}_2 \dot{x}_3 c^{-2} & 1 - (\dot{x}_1^2 + \dot{x}_2^2)c^{-2} \end{vmatrix},$$

$$m_0 \ddot{x}_2 = (1 - v^2 c^{-2})^{-1/2} \begin{vmatrix} 1 - (\dot{x}_2^2 + \dot{x}_3^2)c^{-2} & X_1 & \dot{x}_1 \dot{x}_3 c^{-2} \\ \dot{x}_1 \dot{x}_2 c^{-2} & X_2 & \dot{x}_2 \dot{x}_3 c^{-2} \\ \dot{x}_1 \dot{x}_3 c^{-2} & X_3 & 1 - (\dot{x}_1^2 + \dot{x}_2^2)c^{-2} \end{vmatrix}, \tag{3}$$

$$m_0 \ddot{x}_3 = (1 - v^2 c^{-2})^{-1/2} \begin{vmatrix} 1 - (\dot{x}_2^2 + \dot{x}_3^2)c^{-2} & \dot{x}_1 \dot{x}_2 c^{-2} & X_1 \\ \dot{x}_1 \dot{x}_2 c^{-2} & 1 - (\dot{x}_1^2 + \dot{x}_3^2)c^{-2} & X_2 \\ \dot{x}_1 \dot{x}_3 c^{-2} & \dot{x}_2 \dot{x}_3 c^{-2} & X_3 \end{vmatrix}.$$

These equations are, of course, the differential equations of motion (1) written in a new, but equivalent, form.

If, at some particular instant, the particle is moving parallel to the x_1-axis, so that $\dot{x}_2 = \dot{x}_3 = 0$, the equations (3) reduce *at that instant* to the forms:

$$\frac{m_0 \ddot{x}_1}{(1 - v^2 c^{-2})^{3/2}} = X_1, \qquad \frac{m_0 \ddot{x}_2}{(1 - v^2 c^{-2})^{1/2}} = X_2, \qquad \frac{m_0 \ddot{x}_3}{(1 - v^2 c^{-2})^{1/2}} = X_3.$$

These equations show that a particle of rest-mass m_0, moving with speed v, responds to a force parallel to the velocity as would a Newtonian particle[5] of mass

$$m_\ell = \frac{m_0}{(1 - v^2 c^{-2})^{3/2}},$$

[4] If this conclusion is not entirely evident, the reader is referred to equations (3), from which the conclusion follows at once.

[5] I.e. an ideal particle which obeys the laws of Newtonian dynamics.

and that the particle responds to a force perpendicular to the velocity as would a Newtonian particle of mass

$$m_t = \frac{m_0}{(1 - v^2 c^{-2})^{1/2}}.$$

For this reason, it was usual in the early work on relativistic dynamics to ascribe two masses to a particle: the *longitudinal mass* m_ℓ, and the *transverse mass* m_t. However, in general this procedure leads only to inconveniences, and it has been almost entirely abandoned.

This concludes our discussion of the elementary differential equations of motion. Without any further general theory of relativistic dynamics it is possible to solve many interesting and important problems. For instance, it can be shown easily that the trajectory of a particle subjected to a force which is constant in magnitude and direction is a catenary (rather than a parabola, which is the curve predicted by Newtonian dynamics).[6] In the following sections we shall discuss some of the less elementary parts of the subject.

III. The Lagrangian Equations

In the foregoing the components of the applied force have been any functions of the coordinates, the components of the velocity, and the time. However, in problems concerning the motion of electrons, and for that matter in many other physical problems also, we are usually concerned with forces of a somewhat special kind. Throughout the remainder of the article we shall assume that the force belongs to this special class.

We consider four given functions of the coordinates and time, namely

$$V(x_1, x_2, x_3, t), \qquad A_n(x_1, x_2, x_3, t), \qquad (n = 1, 2, 3),$$

and we assume that the components of the force are given by the formulae

$$X_1 = -\frac{\partial V}{\partial x_1} - \frac{\partial A_1}{\partial t} + \dot{x}_2 \left[\frac{\partial A_2}{\partial x_1} - \frac{\partial A_1}{\partial x_2} \right] - \dot{x}_3 \left[\frac{\partial A_1}{\partial x_3} - \frac{\partial A_3}{\partial x_1} \right],$$

$$X_2 = -\frac{\partial V}{\partial x_2} - \frac{\partial A_2}{\partial t} + \dot{x}_3 \left[\frac{\partial A_3}{\partial x_2} - \frac{\partial A_2}{\partial x_3} \right] - \dot{x}_1 \left[\frac{\partial A_2}{\partial x_1} - \frac{\partial A_1}{\partial x_2} \right], \qquad (4)$$

$$X_3 = -\frac{\partial V}{\partial x_3} - \frac{\partial A_3}{\partial t} + \dot{x}_1 \left[\frac{\partial A_1}{\partial x_3} - \frac{\partial A_3}{\partial x_1} \right] - \dot{x}_2 \left[\frac{\partial A_3}{\partial x_2} - \frac{\partial A_2}{\partial x_3} \right].$$

Let us suppose, for purposes of illustration, that we are considering the motion of an electron. Then the physical interpretation of our assumption

---

[6] L. A. MacColl, *American Mathematical Monthly*, Vol. 45 (1938), pp. 669–676.

concerning the force is the following. $V(x_1, x_2, x_3, t)$ is the potential energy of the electron in an electromagnetic field; that is

$$V(x_1, x_2, x_3, t) = - \epsilon\varphi(x_1, x_2, x_3, t),$$

where ϵ is the absolute value of the electronic charge, and $\varphi(x_1, x_2, x_3, t)$ is the scalar potential of the field. The functions $A_n(x_1, x_2, x_3, t)$ are related to the components $a_n(x_1, x_2, x_3, t)$ of the vector potential of the field by the equations

$$A_n(x_1, x_2, x_3, t) = - \epsilon a_n(x_1, x_2, x_3, t).$$

The terms $-\partial A_n/\partial t$ are $-\epsilon$ times the contributions of the vector potential to the components of the electric force. The quantity $\partial A_3/\partial x_2 - \partial A_2/\partial x_3$ is $-\epsilon B_1$, where B_1 is the x_1-component of the magnetic induction; and similarly for the quantities $\partial A_1/\partial x_3 - \partial A_3/\partial x_1$ and $\partial A_2/\partial x_1 - \partial A_1/\partial x_2$.* In other cases also, equations (4), which may degenerate considerably, can be interpreted without difficulty.

Now we define a function $L(x_1, x_2, x_3, \dot{x}_1, \dot{x}_2, \dot{x}_3, t)$ of the coordinates, the components of the velocity, and the time, as follows:

$$L = - m_0c^2(1 - v^2c^{-2})^{1/2} - V + \dot{x}_1A_1 + \dot{x}_2A_2 + \dot{x}_3A_3. \tag{5}$$

We call this the Lagrangian function.

We write the equations

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{x}_n} - \frac{\partial L}{\partial x_n} = 0, \qquad (n = 1, 2, 3), \tag{6}$$

carry out the indicated differentiations, and readily verify that the resulting equations are identical with those obtained by substituting the expressions (4) in equations (1). Hence, equations (6) are merely a form of the differential equations of motion. We call equations (6) the Lagrangian equations. The chief importance of these equations is due to the ease with which they enable us to use coordinate systems which are not rectangular. This will be discussed in the final section.

In the Newtonian case, i.e. the case in which the speed of the particle is small compared with the speed of light, the Lagrangian function reduces approximately to the form

$$L = -m_0 c^2 + \frac{m_0}{2}(\dot{x}_1^2 + \dot{x}_2^2 + \dot{x}_3^2) - V + \dot{x}_1 A_1 + \dot{x}_2 A_2 + \dot{x}_3 A_3. \tag{5'}$$

---

* These relations between the A's and the components of the vector potential, and between the partial derivatives of the A's and the components of the magnetic induction, are based upon the use of the M.K.S. system of units. If we measure the electromagnetic quantities in other units, certain constant proportionality factors may appear in the relations.

If we employ the function (5') in equations (6), we do indeed get the Newtonian differential equations. Since the constant term $-m_0c^2$ is of no effect in the formation of the differential equations of motion, it is ordinarily omitted in writing the Newtonian form of the Lagrangian function.

IV. HAMILTON'S CANONICAL EQUATIONS

Let us write

$$p_n + A_n = \pi_n. \tag{7}$$

Solving equations (2) for $\dot{x}_1, \dot{x}_2, \dot{x}_3$, we get the result

$$
\begin{aligned}
\dot{x}_n &= cp_n[m_0^2 c^2 + p_1^2 + p_2^2 + p_3^2]^{-1/2} \\
&= c(\pi_n - A_n)[m_0^2 c^2 + (\pi_1 - A_1)^2 + (\pi_2 - A_2)^2 + (\pi_3 - A_3)^2]^{-1/2}.
\end{aligned} \tag{8}
$$

Also, it is readily seen that the differential equations (1) can be written, with the aid of equations (7) and (8), in the form

$$
\begin{aligned}
\dot{\pi}_n &= -\frac{\partial V}{\partial x_n} + \dot{x}_1 \frac{\partial A_1}{\partial x_n} + \dot{x}_2 \frac{\partial A_2}{\partial x_n} + \dot{x}_3 \frac{\partial A_3}{\partial x_n} \\
&= -\frac{\partial V}{\partial x_n} - c\frac{\partial}{\partial x_n}[m_0^2 c^2 + (\pi_1 - A_1)^2 + (\pi_2 - A_2)^2 + (\pi_3 - A_3)^2]^{1/2}.
\end{aligned} \tag{9}
$$

Now let us define a function $H(x_1, x_2, x_3, \pi_1, \pi_2, \pi_3, t)$ as follows:

$$H = c[m_0^2 c^2 + (\pi_1 - A_1)^2 + (\pi_2 - A_2)^2 + (\pi_3 - A_3)^2]^{1/2} + V. \tag{10}$$

Then equations (8) take the forms

$$\dot{x}_n = \frac{\partial H}{\partial \pi_n}, \tag{11}$$

and equations (9) take the forms

$$\dot{\pi}_n = -\frac{\partial H}{\partial x_n}. \tag{12}$$

The function H is called the Hamiltonian function. The six equations (11) and (12), which are equivalent to the three equations (1), are called Hamilton's canonical equations of motion. These equations are of great importance in all of the deeper theoretical work in dynamics.

An easy calculation shows that we have the identity

$$H + L = \pi_1 \dot{x}_1 + \pi_2 \dot{x}_2 + \pi_3 \dot{x}_3. \tag{13}$$

In the Newtonian case the Hamiltonian function given by (10) reduces approximately to the form

$$H = m_0 c^2 + \frac{1}{2m_0} [(\pi_1 - A_1)^2 + (\pi_2 - A_2)^2 + (\pi_3 - A_3)^2] + V. \quad (10')$$

The equations (11) and (12), with H given by (10'), are equivalent to the Newtonian differential equations of motion (1'). Here again the constant term $m_0 c^2$ is of no effect, and it is ordinarily omitted in writing the Newtonian form of the function H. The Newtonian forms of the functions H and L satisfy the identity (13), whether or not the constant terms $m_0 c^2$ and $- m_0 c^2$ are included.

V. Static Fields of Force: The Energy Integral; Natural Families of Trajectories

By equations (11) and (12), we have the relation

$$\begin{aligned}
\frac{dH}{dt} &= \frac{\partial H}{\partial t} + \sum_{n=1}^{3} \left[\frac{\partial H}{\partial x_n} \dot{x}_n + \frac{\partial H}{\partial \pi_n} \dot{\pi}_n \right] \\
&= \frac{\partial H}{\partial t} + \sum_{n=1}^{3} \left[\frac{\partial H}{\partial x_n} \frac{\partial H}{\partial \pi_n} - \frac{\partial H}{\partial \pi_n} \frac{\partial H}{\partial x_n} \right] = \frac{\partial H}{\partial t} . \quad (14)
\end{aligned}$$

In particular, if no one of the functions V, A_1, A_2, A_3 involves the time explicitly, we have $dH/dt = 0$, so that the value of H remains constant during the motion of the particle. That is, under the condition stated we have

$$m_0 c^2 [1 - v^2 c^{-2}]^{-1/2} + V(x_1, x_2, x_3) = \text{constant}. \quad (15)$$

In the Newtonian case equation (15) reduces approximately to the form

$$m_0 c^2 + \frac{m_0}{2} v^2 + V(x_1, x_2, x_3) = \text{constant},$$

which is equivalent to the equation

$$\frac{m_0}{2} v^2 + V(x_1, x_2, x_3) = \text{constant}. \quad (15')$$

It is well known that this equation is a consequence of the Newtonian differential equations of motion.

The left-hand member of equation (15') is the energy of the particle in

Newtonian dynamics, the first and second terms being the kinetic energy and the potential energy, respectively. The equation itself is called the *energy integral*.[7] Similarly, we call (15) the energy integral in relativistic dynamics, and we call the expression

$$m_0 c^2 [1 - v^2 c^{-2}]^{-1/2} + V$$

the *relativistic energy*. This energy is the sum of three parts: the *proper energy $m_0 c^2$*, the *relativistic kinetic energy*

$$m_0 c^2 [1 - v^2 c^{-2}]^{-1/2} - m_0 c^2,$$

and the *potential energy V*.

The totality of possible trajectories of a particle in a static field of force forms a five-parameter family. We now see that if the field of force is static and of the kind we are considering now, the five-parameter family of curves consists of ∞^1 four-parameter subfamilies, each of which corresponds to a different value of the energy of the particle. Each of these four-parameter subfamilies is called a *natural family* of trajectories. We proceed to derive the differential equations defining a natural family.

If the constant in the right-hand member of equation (15) is denoted by the symbol E, we have the relation

$$\dot{x}_1 [1 + x_2'^2 + x_3'^2]^{1/2} = c[1 - m_0^2 c^4 (E - V)^{-2}]^{1/2}, \tag{16}$$

where

$$x_2' = dx_2/dx_1, \qquad x_3' = dx_3/dx_1.$$

Hence,

$$dt = c^{-1}[1 + x_2'^2 + x_3'^2]^{1/2}[1 - m_0^2 c^4 (E - V)^{-2}]^{-1/2} \, dx_1.$$

From this, and the two equations

$$\frac{d}{dt} \frac{m_0 \dot{x}_2}{(1 - v^2 c^{-2})^{1/2}} = -\frac{\partial V}{\partial x_2} + \dot{x}_3 \left[\frac{\partial A_3}{\partial x_2} - \frac{\partial A_2}{\partial x_3} \right] - \dot{x}_1 \left[\frac{\partial A_2}{\partial x_1} - \frac{\partial A_1}{\partial x_2} \right],$$

$$\frac{d}{dt} \frac{m_0 \dot{x}_3}{(1 - v^2 c^{-2})^{1/2}} = -\frac{\partial V}{\partial x_3} + \dot{x}_1 \left[\frac{\partial A_1}{\partial x_3} - \frac{\partial A_3}{\partial x_1} \right] - \dot{x}_2 \left[\frac{\partial A_3}{\partial x_2} - \frac{\partial A_2}{\partial x_3} \right],$$

it follows that we have the following system of differential equations defining the natural family of trajectories corresponding to the total energy E:

---

[7] In the theory of differential equations, an equation relating the unknowns involved in a system of differential equations, their derivatives of orders less than the highest orders appearing in the system, the independent variable, and one or more arbitrary constants, is called an integral of the system of differential equations.

$$[1 + x_2'^2 + x_3'^2]^{-1/2} \frac{d}{dx_1}\left(x_2'\left[\frac{(E - V)^2 - m_0^2 c^4}{1 + x_2'^2 + x_3'^2} \right]^{1/2} \right)$$

$$= \frac{\partial}{\partial x_2}[(E - V)^2 - m_0^2 c^4]^{1/2}$$

$$+ c[1 + x_2'^2 + x_3'^2]^{-1/2}\left(x_3'\left[\frac{\partial A_3}{\partial x_2} - \frac{\partial A_2}{\partial x_3} \right] - \left[\frac{\partial A_2}{\partial x_1} - \frac{\partial A_1}{\partial x_2} \right] \right),$$

$$[1 + x_2'^2 + x_3'^2]^{-1/2} \frac{d}{dx_1}\left(x_3'\left[\frac{(E - V)^2 - m_0^2 c^4}{1 + x_2'^2 + x_3'^2} \right]^{1/2} \right) \tag{17}$$

$$= \frac{\partial}{\partial x_3}[(E - V)^2 - m_0^2 c^4]^{1/2}$$

$$+ c[1 + x_2'^2 + x_3'^2]^{-1/2}\left(\left[\frac{\partial A_1}{\partial x_3} - \frac{\partial A_3}{\partial x_1} \right] - x_2'\left[\frac{\partial A_3}{\partial x_2} - \frac{\partial A_2}{\partial x_3} \right] \right).$$

The equations which correspond to (17) in the Newtonian case are most readily obtained by going back to the Newtonian differential equations of motion and employing the integral

$$m_0 v^2/2 + V = E.$$

An easy calculation, which is entirely parallel to the foregoing, gives us the following system of equations:

$$[1 + x_2'^2 + x_3'^2]^{-1/2} \frac{d}{dx_1}\left(x_2'\left[\frac{E - V}{1 + x_2'^2 + x_3'^2} \right]^{1/2} \right) = \frac{\partial}{\partial x_2}(E - V)^{1/2}$$

$$+ [2m_0(1 + x_2'^2 + x_3'^2)]^{-1/2}\left(x_3'\left[\frac{\partial A_3}{\partial x_2} - \frac{\partial A_2}{\partial x_3} \right] - \left[\frac{\partial A_2}{\partial x_1} - \frac{\partial A_1}{\partial x_2} \right] \right),$$

$$[1 + x_2'^2 + x_3'^2]^{-1/2} \frac{d}{dx_1}\left(x_3'\left[\frac{E - V}{1 + x_2'^2 + x_3'^2} \right]^{1/2} \right) = \frac{\partial}{\partial x_3}(E - V)^{1/2} \tag{17'}$$

$$+ [2m_0(1 + x_2'^2 + x_3'^2)]^{-1/2}\left(\left[\frac{\partial A_1}{\partial x_3} - \frac{\partial A_3}{\partial x_1} \right] - x_2'\left[\frac{\partial A_3}{\partial x_2} - \frac{\partial A_2}{\partial x_3} \right] \right).$$

On comparing the systems of equations (17) and (17'), we get the following useful theorem.

If the constants E, E^*, m_0, m_0^*, k, and the functions (of x_1, x_2, x_3,) V, A_1, A_2, A_3, V^*, A_1^*, A_2^*, A_3^* are such that we have identically

$$(E - V)^2 - m_0^2 c^4 = k^2(E^* - V^*),$$

$$\frac{\partial A_3}{\partial x_2} - \frac{\partial A_2}{\partial x_3} = \frac{k}{c(2m_0^*)^{1/2}}\left[\frac{\partial A_3^*}{\partial x_2} - \frac{\partial A_2^*}{\partial x_3} \right],$$

$$\frac{\partial A_2}{\partial x_1} - \frac{\partial A_1}{\partial x_2} = \frac{k}{c(2m_0{}^*)^{1/2}} \left[\frac{\partial A_2^*}{\partial x_1} - \frac{\partial A_1^*}{\partial x_2} \right],$$

$$\frac{\partial A_1}{\partial x_3} - \frac{\partial A_3}{\partial x_1} = \frac{k}{c(2m_0{}^*)^{1/2}} \left[\frac{\partial A_1^*}{\partial x_3} - \frac{\partial A_3^*}{\partial x_1} \right],$$

the natural family of trajectories of a relativistic particle[8] *(of rest-mass m_0) moving with relativistic total energy E in the field of force derived from the functions V, A_1, A_2, A_3 is identical with the natural family of trajectories of a Newtonian particle (of mass $m_0{}^*$) moving with Newtonian energy E^* in the field of force derived from the functions V^*, $A_1{}^*$, $A_2{}^*$, $A_3{}^*$.*

In particular, the conditions of the theorem are satisfied if

$$k = c(2m_0)^{1/2}, \qquad E^* = c^{-2}(2m_0)^{-1}(E^2 - m_0^2 c^4), \qquad m_0{}^* = m_0,$$

$$V = V^* = 0, \qquad A_1{}^* = A_1, \qquad A_2{}^* = A_2, \qquad A_3{}^* = A_3.$$

Hence, we have the corollary:

In the case of an electrified particle moving in any static magnetic field the natural family of trajectories corresponding to any value of the energy given by relativistic (Newtonian) dynamics is identical with the natural family of trajectories corresponding to a certain other value of the energy given by Newtonian (relativistic) dynamics.

The equation

$$E^* = c^{-2}(2m_0)^{-1}(E^2 - m_0^2 c^4)$$

establishes a one-to-one correspondence between the physically significant ($E \geqq m_0 c^2$ and $E^* \geqq 0$) values of the relativistic energy E and the Newtonian energy E^*. From this fact and the preceding corollary we get the following further result:

In the case of an electrified particle moving in any static magnetic field the total five-parameter family of trajectories given by relativistic dynamics is identical with that given by Newtonian dynamics.

Of course, these peculiar properties of motion of an electrified particle moving in a static magnetic field are explained physically by the fact that the magnetic forces do no work, so that the speed of the particle, and consequently also its mass, remain constant during the motion.

VI. Some Formulae from the Calculus of Variations

This section is devoted to the derivation of some formulae from the Calculus of Variations which will be needed in the further discussion of the

---

[8] I.e. a particle obeying the laws of relativistic dynamics.

dynamics of a particle. All constants, variables, and functions considered here are understood to be real.[9]

Let $F(t, x, y, z, p, q, r)$ be a function of the seven arguments indicated,[10] which, together with all of its partial derivatives of the first three orders, is continuous in a region R defined as follows:

$$a_1 < t < a_2,$$

$$b_1 < x < b_2,$$

$$R: \quad c_1 < y < c_2,$$

$$d_1 < z < d_2,$$

$$p, q, \text{ and } r \text{ unrestricted,}$$

the a's, b's, c's, and d's, being constants.

Let $x(t)$, $y(t)$, $z(t)$, $\varphi(t)$, $\psi(t)$, and $\omega(t)$ be continuous functions with continuous first derivatives, and let ϵ, η, and θ be parameters, independent of t, such that we have the relations

$$b_1 < x(t) + \epsilon\varphi(t) < b_2,$$

$$c_1 < y(t) + \eta\psi(t) < c_2, \quad (a_1 < t < a_2).$$

$$d_1 < z(t) + \theta\omega(t) < d_2,$$

Let T_1 and T_2 be constants, and let t_1 and t_2 be parameters, such that

$$a_1 < T_1 + t_1 < T_2 + t_2 < a_2.$$

We now consider the integral

$$I(\epsilon, \eta, \theta, t_1, t_2)$$

$$= \int_{T_1+t_1}^{T_2+t_2} F(t, x + \epsilon\varphi, y + \eta\psi, z + \theta\omega, x' + \epsilon\varphi', y' + \eta\psi', z' + \theta\omega') \, dt.$$

It can be shown without difficulty that the integral exists and is a differentiable function of ϵ, η, θ, t_1, t_2. We are interested in formulae giving the values of $\partial I/\partial\epsilon$, $\partial I/\partial\eta$, $\partial I/\partial\theta$, $\partial I/\partial t_1$, $\partial I/\partial t_2$ at the point $\epsilon = \eta = \theta = t_1 = t_2 = 0$.

---

[9] Since this section is purely mathematical, the constants, variables, and functions do not necessarily have any special physical significance.

[10] We treat the case of a function of seven arguments in order to fix the ideas, and because this is a case we shall meet in Section VII. However, the discussion applies essentially to other cases as well. In particular, in Section VII we shall also deal with a case in which F has only five arguments, z and r being absent.

By a well known theorem concerning the differentiation of definite integrals with respect to parameters,[11] we have

$$\frac{\partial I}{\partial \epsilon} = \int_{T_1+t_1}^{T_2+t_2} \left[\varphi \frac{\partial}{\partial (x + \epsilon\varphi)} + \varphi' \frac{\partial}{\partial (x' + \epsilon\varphi')} \right] F(t, x + \epsilon\varphi, \cdots, z' + \theta\omega') \, dt,$$

$$\frac{\partial I}{\partial t_1} = -F[T_1 + t_1, x(T_1 + t_1) + \epsilon\varphi(T_1 + t_1), \cdots, z'(T_1 + t_1) + \theta\omega'(T_1 + t_1)],$$

$$\frac{\partial I}{\partial t_2} = F[T_2 + t_2, x(T_2 + t_2) + \epsilon\varphi(T_2 + t_2), \cdots, z'(T_2 + t_2) + \theta\omega'(T_2 + t_2)].$$

The formulae for $\partial I/\partial \eta$ and $\partial I/\partial \theta$ are similar to that for $\partial I/\partial \epsilon$, and need not be written down.

In particular, if $[\partial I/\partial \epsilon]_0$, etc. denote the values of the derivatives at the point $\epsilon = \eta = \theta = t_1 = t_2 = 0$, we have

$$\left[\frac{\partial I}{\partial \epsilon} \right]_0 = \int_{T_1}^{T_2} \left[\varphi \frac{\partial}{\partial x} + \varphi' \frac{\partial}{\partial x'} \right] F(t, x, \cdots, z') \, dt,$$

$$\left[\frac{\partial I}{\partial \eta} \right]_0 = \int_{T_1}^{T_2} \left[\psi \frac{\partial}{\partial y} + \psi' \frac{\partial}{\partial y'} \right] F(t, x, \cdots, z') \, dt,$$

$$\left[\frac{\partial I}{\partial \theta} \right]_0 = \int_{T_1}^{T_2} \left[\omega \frac{\partial}{\partial z} + \omega' \frac{\partial}{\partial z'} \right] F(t, x, \cdots, z') \, dt, \qquad (18)$$

$$\left[\frac{\partial I}{\partial t_1} \right]_0 = -F[T_1, x(T_1), \cdots, z'(T_1)],$$

$$\left[\frac{\partial I}{\partial t_2} \right]_0 = F[T_2, x(T_2), \cdots, z'(T_2)].$$

The first three of equations (18) can be transformed to advantage, as follows. Integrating by parts, we obtain the formula

$$\int_{T_1}^{T_2} \varphi' \frac{\partial}{\partial x'} F(t, x, \cdots, z') dt = \left[\varphi \frac{\partial}{\partial x'} F(t, x, \cdots, z') \right]_{T_1}^{T_2}$$
$$- \int_{T_1}^{T_2} \varphi \frac{d}{dt} \frac{\partial}{\partial x'} F(t, x, \cdots, z') \, dt,$$

and similar formulae for the integrals

$$\int_{T_1}^{T_2} \psi' \frac{\partial}{\partial y'} F(t, x, \cdots, z') \, dt$$

and

$$\int_{T_1}^{T_2} \omega' \frac{\partial}{\partial z'} F(t, x, \cdots, z') \, dt.$$

---

[11] The theorem is given, often in the form of two separate theorems, in most works on Advanced Calculus and the Theory of Functions of Real Variables. See the bibliography.

It follows, therefore, that we have

$$
\left[\frac{\partial I}{\partial \epsilon}\right]_0 = \left[\varphi \frac{\partial}{\partial x'} F(t, x, \cdots, z')\right]_{T_1}^{T_2}
$$
$$
+ \int_{T_1}^{T_2} \varphi(t) \left[\frac{\partial}{\partial x} - \frac{d}{dt}\frac{\partial}{\partial x'}\right] F(t, x, \cdots, z') \, dt,
$$

$$
\left[\frac{\partial I}{\partial \eta}\right]_0 = \left[\psi \frac{\partial}{\partial y'} F(t, x, \cdots, z')\right]_{T_1}^{T_2}
$$
$$
+ \int_{T_1}^{T_2} \psi(t) \left[\frac{\partial}{\partial y} - \frac{d}{dt}\frac{\partial}{\partial y'}\right] F(t, x, \cdots, z') \, dt, \quad (19)
$$

$$
\left[\frac{\partial I}{d\theta}\right]_0 = \left[\omega \frac{\partial}{\partial z'} F(t, x, \cdots, z')\right]_{T_1}^{T_2}
$$
$$
+ \int_{T_1}^{T_2} \omega(t) \left[\frac{\partial}{\partial z} - \frac{d}{dt}\frac{\partial}{\partial z'}\right] F(t, x, \cdots, z') \, dt.
$$

An important special case is that in which t_1 and t_2 are zero (so that the limits of integration are fixed), and
$$
\varphi(T_1) = \varphi(T_2) = \psi(T_1) = \psi(T_2) = \omega(T_1) = \omega(T_2) = 0.
$$
In this case we have in general

$$
I(\epsilon, \eta, \theta, 0, 0) - I(0, 0, 0, 0, 0) = \epsilon \int_{T_1}^{T_2} \varphi(t) \left[\frac{\partial F}{\partial x} - \frac{d}{dt}\frac{\partial F}{\partial x'}\right] dt
$$
$$
+ \eta \int_{T_1}^{T_2} \psi(t) \left[\frac{\partial F}{\partial y} - \frac{d}{dt}\frac{\partial F}{\partial y'}\right] dt + \theta \int_{T_1}^{T_2} \omega(t) \left[\frac{\partial F}{\partial z} - \frac{d}{dt}\frac{\partial F}{\partial z'}\right] dt
$$
$$
+ o(\epsilon, \eta, \theta),
$$

where $o(\epsilon, \eta, \theta)$ denotes a term, the exact form of which is unimportant, *which is such that the expression*

$$
\frac{o(\epsilon, \eta, \theta)}{|\epsilon| + |\eta| + |\theta|}
$$

approaches the limit zero as ϵ, η, and θ tend simultaneously toward zero.

In particular, if the functions $x(t)$, $y(t)$, $z(t)$ satisfy the system of differential equations

$$
\frac{d}{dt}\frac{\partial F}{dx'} - \frac{\partial F}{\partial x} = 0, \qquad \frac{d}{dt}\frac{\partial F}{\partial y'} - \frac{\partial F}{\partial y} = 0, \qquad \frac{d}{dt}\frac{\partial F}{dz'} - \frac{\partial F}{\partial z} = 0, \quad (20)
$$

we have (for all choices of the functions φ, ψ, ω subject to the conditions stated)

$$
I(\epsilon, \eta, \theta, 0, 0) - I(0, 0, 0, 0, 0) = o(\epsilon, \eta, \theta). \quad (21)
$$

Also, it can be shown without difficulty that in order that we have (21), for all such choices of φ, ψ, ω, it is necessary that x, y, and z, satisfy the equations (20).*

The last result can be stated in the following summary, and not quite explicit, form: If, and only if, the functions $x(t)$, $y(t)$, $z(t)$ satisfy equations (20), the integral

$$\int_{T_1}^{T_2} F(t, x, y, z, x', y', z')\, dt \tag{22}$$

is stationary with respect to infinitesimal variations of the functions $x(t)$, $y(t)$, $z(t)$ which leave the terminal values unaltered.

The problem of finding functions which render the values of definite integrals stationary is the chief subject of the Calculus of Variations.

The equations (20) are called the Eulerian equations of the Calculus of Variations problem of making the value of the integral (22) stationary, or, as we usually say, of maximizing or minimizing the integral.

VII. Hamilton's Principle and the Principle of Least Action

We immediately recognize equations (6) as the Eulerian equations of a problem in the Calculus of Variations. Thus we have the following principle (*Hamilton's principle*):

The particle moves, under forces of the type (4), *so that the value of the integral*

$$\int_{t_1}^{t_2} L\, dt,$$

with t_1 and t_2 held fixed, is stationary with respect to infinitesimal variations of the functions $x_n(t)$ which leave the initial and final points unaltered.

The precise meaning of this is determined by the discussion given in Section VI.

Hamilton's principle leads to the relativistic or Newtonian differential equations of motion, according as we use in it the function L given by (5) or by (5′).

A little inspection suffices to show that the system of equations (17) is also the system of Eulerian equations of a problem in the Calculus of

---

* In brief, suppose that $\dfrac{d}{dt}\dfrac{\partial F}{\partial x'} - \dfrac{\partial F}{\partial x}$ were not zero for some value of t. Then if we should choose a function $\varphi(t)$ which was (say) positive in the neighborhood of that value, and zero elsewhere, the integral

$$\int_{T_1}^{T_2} \varphi(t) \left[\frac{\partial F}{\partial x} - \frac{d}{dt}\frac{\partial F}{\partial x'} \right] dt$$

would have a value other than zero. We shall not give the actual proof here; it is to be found in the works on the Calculus of Variations cited in the bibliography.

Variations.   Thus we get the so-called *principle of least action*, which can be stated as follows:

The particle moves, in a static field of force of the type (4), *and with the prescribed total energy E, in such a curve that the value of the integral*

$$\int_{(x_1)_1}^{(x_1)_2} ([1 + x_2'^2 + x_3'^2]^{1/2}[(E - V)^2 c^{-2} - m_0^2 c^2]^{1/2} + A_1 + A_2 x_2' + A_3 x_3') \, dx_1,$$

with the limits of integration held fixed, is stationary with respect to infinitesimal variations of the trajectory which leave the end points unaltered.

We have a precisely similar principle in Newtonian dynamics, but here the integral in question is

$$\int_{(x_1)_1}^{(x_1)_2} ([1 + x_2'^2 + x_3'^2]^{1/2}[2m_0(E - V)]^{1/2} + A_1 + A_2 x_2' + A_3 x_3') \, dx_1.$$

The last two integrals can be written more symmetrically, but not quite so explicitly, as follows:

$$\int_{P_1}^{P_2} \left([(E - V)^2 c^{-2} - m_0^2 c^2]^{1/2} + A_1 \frac{dx_1}{ds} + A_2 \frac{dx_2}{ds} + A_3 \frac{dx_3}{ds} \right) ds,$$

$$\int_{P_1}^{P_2} \left([2m_0(E - V)]^{1/2} + A_1 \frac{dx_1}{ds} + A_2 \frac{dx_2}{ds} + A_3 \frac{dx_3}{ds} \right) ds,$$

where P_1 and P_2 denote the end points of the trajectory, and $ds^2 = dx_1^2 + dx_2^2 + dx_3^2.$

VIII. The Hamilton-Jacobi Theory

Let us write

$$W = \int_{t_1}^{t_2} L[x_1(t), x_2(t), x_3(t), x_1'(t), x_2'(t), x_3'(t), t] \, dt. \tag{23}$$

We have already studied the variation of W when t_1 and t_2 are held fixed, and the functions $x_n(t)$ are varied in such a way that the terminal values are unaltered; and we have shown that under these circumstances the variation of W vanishes, to the first order of small quantities, in the natural motion.[12] In the following we shall study the variation of W under some other conditions.

Specifically, we shall study the quantity ΔW defined by equation (23) and the equation

$$W + \Delta W = \int_{t_1 + \Delta t_1}^{t_2 + \Delta t_2} L[x_1(t) + \xi_1(t), \cdots, x_3'(t) + \xi_3'(t), t] \, dt,$$

---

[12] I.e. a motion satisfying equations (1).

where the functions $x_n(t)$ represent a natural motion, the $\xi_n(t)$ are small functions, and Δt_1 and Δt_2 are small parameters.

It follows from the results of Section VI that we have (to within terms of the second order in small quantities)[13]

$$
\begin{aligned}
\Delta W = \quad & \Delta t_2\, L[x_1(t_2), \cdots, x_3'(t_2), t_2] \\
& - \Delta t_1\, L[x_1(t_1), \cdots, x_3'(t_1), t_1] \\
& + \sum_{n=1}^{3} \left[\frac{\partial L}{\partial x_n'}\right]_{t=t_2} \xi_n(t_2) - \sum_{n=1}^{3} \left[\frac{\partial L}{\partial x_n'}\right]_{t=t_1} \xi_n(t_1) \\
= \quad & \Delta t_2\, L[x_1(t_2), \cdots, x_3'(t_2), t_2] \\
& - \Delta t_1\, L[x_1(t_1), \cdots, x_3'(t_1), t_1] \\
& + \sum_{n=1}^{3} [\pi_n(t_2)\xi_n(t_2) - \pi_n(t_1)\xi_n(t_1)].
\end{aligned}
$$

Let us write

$$
(\Delta x_n)_2 = x_n(t_2 + \Delta t_2) + \xi_n(t_2 + \Delta t_2) - x_n(t_2) = \xi_n(t_2) + x_n'(t_2)\,\Delta t_2,
$$
$$
(\Delta x_n)_1 = x_n(t_1 + \Delta t_1) + \xi_n(t_1 + \Delta t_1) - x_n(t_1) = \xi_n(t_1) + x_n'(t_1)\Delta t_1,
$$

so that $(\Delta x_1)_2$, $(\Delta x_2)_2$, $(\Delta x_3)_2$ are the coordinate differences of the terminal points of the varied and unvaried curves, and similarly $(\Delta x_1)_1$, $(\Delta x_2)_1$, $(\Delta x_3)_1$ are the coordinate differences of the initial points. Then we have the formula

$$
\begin{aligned}
\Delta W = \quad & \left(L[x_1(t_2), \cdots] - \sum_{n=1}^{3} \pi_n(t_2)x_n'(t_2) \right) \Delta t_2 \\
& - \left(L[x_1(t_1), \cdots] - \sum_{n=1}^{3} \pi_n(t_1)x_n'(t_1) \right) \Delta t_1 \\
& + \sum_{n=1}^{3} [\pi_n(t_2)(\Delta x_n)_2 - \pi_n(t_1)(\Delta x_n)_1],
\end{aligned}
$$

which, by equation (13), can be written in the form

$$
\Delta W = -H[x_1(t_2), \cdots]\Delta t_2 + H[x_1(t_1), \cdots]\Delta t_1
$$
$$
+ \sum_{n=1}^{3} [\pi_n(t_2)(\Delta x_n)_2 - \pi_n(t_1)(\Delta x_n)_1]. \quad (24)
$$

Now, the integration in (23) being taken over a natural motion of the particle, the value of W depends upon the initial instant, the initial coordi-

---

[13] This is also the sense in which the following equations are to be understood.

nates, the initial components of velocity, and the final instant. It is necessary now to consider W as depending upon the following equivalent set of eight variables: the initial and final instants t_1 and t_2, the coordinates (x_{11}, x_{21}, x_{31}) of the initial point, and the coordinates (x_{12}, x_{22}, x_{32}) of the final point. Regarding W in this manner, we at once obtain the following relations from equation (24)

$$\frac{\partial W}{\partial t_2} = -H_2, \qquad \frac{\partial W}{\partial x_{n2}} = \pi_{n2}, \tag{25}$$

$$\frac{\partial W}{\partial t_1} = H_1, \qquad \frac{\partial W}{\partial x_{n1}} = -\pi_{n1}, \tag{26}$$

where H_2 denotes $H[x_1(t_2), \cdots, \pi_1(t_2), \cdots, t_2]$ and H_1 denotes $H[x_1(t_1), \cdots, \pi_1(t_1), \cdots, t_1]$.

Let us now consider the partial differential equation

$$\frac{\partial W}{\partial t} + H(x_1, x_2, x_3, \partial W/\partial x_1 \ \partial W/\partial x_2, \partial W/\partial x_3, t) = 0. \tag{27}$$

The preceding work shows that the function W we have been considering (with $x_{11}, x_{21}, x_{31}, t_1$ regarded as parameters, and with the symbols $x_{12}, x_{22}, x_{32}, t_2$ replaced by x_1, x_2, x_3, t respectively) is a *particular* solution of this equation. We shall show that the *complete* solution of this equation possesses remarkable properties in connection with dynamical problems.

The complete solution of equation (27) is a function of x_1, x_2, x_3, t, and of four arbitrary constants, of which one is merely additive, and can be neglected for our purposes. Let the solution be written

$$W = W(x_1, x_2, x_3, t, \alpha_1, \alpha_2, \alpha_3),$$

where the α's are the three essential arbitrary constants.

We write the equations

$$\frac{\partial W}{\partial \alpha_n} = \beta_n, \tag{28}$$

where the β's are further arbitrary constants. These equations implicitly determine the x's as functions of t and the six arbitrary constants $\alpha_1, \cdots, \beta_3$.

We also write the equations

$$\frac{\partial W}{\partial x_n} = \pi_n. \tag{29}$$

These equations determine three functions π_n of the x's, the α's, and t. In virtue of equations (28), the π's are ultimately functions of t, the α's, and the β's.

There is no reason to foresee *a priori* that the functions $x_1(t, \alpha_1, \cdots, \beta_3)$, $\cdots, \pi_3(t, \alpha_1, \cdots, \beta_3)$ determined in this way, by means of the complete solution of equation (27), satisfy the differential equations of motion (11) and (12). Nevertheless, they actually do satisfy those equations, as we proceed to show.

By equations (28), we have the relations

$$0 = \frac{d\beta_n}{dt} = \frac{\partial^2 W}{\partial \alpha_n \partial t} + \sum_{m=1}^{3} \frac{\partial^2 W}{\partial \alpha_n \partial x_m} \dot{x}_m . \tag{30}$$

On the other hand, by (27) and (29), we have[14]

$$0 = \frac{\partial}{\partial \alpha_n} \left[\frac{\partial W}{\partial t} + H(x_1, x_2, x_3, \pi_1, \pi_2, \pi_3, t) \right]$$

$$= \frac{\partial^2 W}{\partial \alpha_n \partial t} + \sum_{m=1}^{3} \frac{\partial H}{\partial \pi_m} \frac{\partial \pi_m}{\partial \alpha_n} = \frac{\partial^2 W}{\partial \alpha_n \partial t} + \sum_{m=1}^{3} \frac{\partial H}{\partial \pi_m} \frac{\partial^2 W}{\partial \alpha_n \partial x_m} . \tag{31}$$

The determinant

$$\begin{vmatrix} \dfrac{\partial^2 W}{\partial \alpha_1 \partial x_1} & \cdots & \dfrac{\partial^2 W}{\partial \alpha_1 \partial x_3} \\ \cdots\cdots & \cdots & \cdots\cdots \\ \dfrac{\partial^2 W}{\partial \alpha_3 \partial x_1} & \cdots & \dfrac{\partial^2 W}{\partial \alpha_3 \partial x_3} \end{vmatrix}$$

is not zero. For if it were, we would have a relation of the form

$$\Phi[\partial W/\partial x_1, \partial W/\partial x_2, \partial W/\partial x_3, x_1, x_2, x_3, t] = 0, \tag{32}$$

independent of the α's. Now equation (32) is obviously distinct from (27), since it does not involve $\partial W/\partial t$. Hence, the vanishing of the determinant would imply that the function $W(x_1, x_2, x_3, t, \alpha_1, \alpha_2, \alpha_3)$ satisfies two distinct partial differential equations of the first order. This, however, is impossible when W is the complete solution of (27); for an essential part of the concept of the complete solution of a differential equation is that the elimination of the arbitrary constants, from the solution and the equations obtained by differentiation, shall result in the given differential equation and no other.

It follows, therefore, from (30) and (31) that

$$\dot{x}_m = \frac{\partial H}{\partial \pi_m} .$$

We also have, by (29),

$$\dot{\pi}_n = \frac{\partial^2 W}{\partial x_n \partial t} + \sum_{m=1}^{3} \frac{\partial^2 W}{\partial x_m \partial x_n} \dot{x}_m = \frac{\partial^2 W}{\partial x_n \partial t} + \sum_{m=1}^{3} \frac{\partial^2 W}{\partial x_m \partial x_n} \frac{\partial H}{\partial \pi_m} . \tag{33}$$

---

[14] Since the function $W(x_1, x_2, x_3, t, \alpha_1, \alpha_2, \alpha_3)$ satisfies equation (27) *identically* in the x's, t, and the α's. This remark applies also in the case of equation (34).

On the other hand, we have

$$0 = \frac{\partial}{\partial x_n}\left[\frac{\partial W}{\partial t} + H\right] = \frac{\partial^2 W}{\partial x_n \partial t} + \frac{\partial H}{\partial x_n} + \sum_{m=1}^{3} \frac{\partial H}{\partial \pi_m} \frac{\partial^2 W}{\partial x_m \partial x_n}. \tag{34}$$

By (33) and (34), we have the second set of canonical equations

$$\dot{\pi}_n = -\frac{\partial H}{\partial x_n}.$$

This completes the demonstration.

If H does not involve the time explicitly, we can write

$$W = S - Et, \tag{35}$$

where E is an arbitrary parameter, and S is a solution of the differential equation

$$H[x_1, x_2, x_3, \partial S/\partial x_1, \partial S/\partial x_2, \partial S/\partial x_3] = E. \tag{36}$$

The complete solution of (36) contains three arbitrary constants (besides the parameter E), of which one is merely additive, and can be neglected. It is easily seen that the solution of the canonical equations determined in the way described above, using the function W given by (35), and treating E as one of the α's, represents a motion of the particle with the total energy E.

All of this theory holds both for the relativistic case and for the Newtonian case, the only difference being in the forms of the differential equations (27) and (36) in the two cases.

IX. Curvilinear Coordinates

In all of the foregoing we have employed rectangular coordinates, because they afford the simplest and most direct expression of the basic physical facts. However, in the solution of particular problems it is often more convenient to use other systems of coordinates. For this reason, we shall now formulate the more important equations in terms of general curvilinear coordinates. In this work, as in all work with general coordinate systems, we shall encounter concepts and relations which can be handled most perspicuously by means of the modern tensor calculus. Actually, the amount of tensor calculus we shall use is very slight, and no extended preliminary discussion is necessary in order to make the formulae intelligible. It will suffice to give occasional explanations of the notation, and of some of the concepts, as we proceed. Further information is to be found in the works cited in the bibliography.

First consider the Lagrangian equations, which, as we have seen, are merely the Eulerian equations which follow from Hamilton's principle.

Now Hamilton's principle expresses a fact concerning the motion of a particle which is, by its very nature, independent of the choice of coordinates. Hence, the Lagrangian equations (6) hold in any coordinate system. However, the form of the function L depends upon the particular coordinate system, and we must discuss the change of the form of the function resulting from a transformation of the coordinate system.

In accordance with the common practice in the tensor calculus, we shall now denote the coordinates by the symbols x^1, x^2, x^3, instead of by the symbols x_1, x_2, x_3.

In rectangular coordinates the differential distance ds between the points (x^1, x^2, x^3) and $(x^1 + dx^1, x^2 + dx^2, x^3 + dx^3)$ is given by the simple formula

$$ds^2 = dx^{1^2} + dx^{2^2} + dx^{3^2},$$

but this is highly special; in general coordinates we have

$$ds^2 = \sum_{m=1}^{3} \sum_{n=1}^{3} g_{mn}(x^1, x^2, x^3) \, dx^m \, dx^n, \tag{37}$$

where the g's are functions which depend upon the particular coordinate system under consideration. It is understood that $g_{mn} = g_{nm}$. Henceforth, we shall write (37) in the form

$$ds^2 = g_{mn}(x^1, x^2, x^3)dx^m dx^n, \tag{38}$$

and we shall observe this general rule throughout: *When the same literal index occurs twice in a term, once as a subscript and once as a superscript, that term is understood to be summed for the three values of the index.*

We now have the result

$$v^2 = [ds/dt]^2 = g_{mn}(x^1, x^2, x^3)\dot{x}^m\dot{x}^n,$$

and

$$m_0c^2(1 - v^2c^{-2})^{1/2} = m_0c^2[1 - c^{-2}g_{mn}\dot{x}^m\dot{x}^n]^{1/2}.$$

The function $V(x^1, x^2, x^3, t)$ is a scalar. That is to say, when the coordinate system is changed, the first three arguments of the function are replaced by their expressions in terms of the new coordinates, and so we obtain a function which is of a new analytical form, but which has the same value as the original function at each point of space.

Now we consider the expression

$$A_1\dot{x}^1 + A_2\dot{x}^2 + A_3\dot{x}^3.$$

In rectangular coordinates this is the scalar product of the vectors (A_1, A_2, A_3) and $(\dot{x}^1, \dot{x}^2, \dot{x}^3)$. The expression retains its form and interpretation

under changes of the coordinate system, provided (as the notation implies) (A_1, A_2, A_3) is treated as a covariant vector.[15]

With these understandings as to the significance of the symbolism, we can now write down the following general expressions for the Lagrangian function L in the relativistic and Newtonian cases, respectively,

$$L = -m_0 c^2 [1 - c^{-2} g_{mn} \dot{x}^m \dot{x}^n]^{1/2} - V + A_m \dot{x}^m ,$$

$$L = -m_0 c^2 + \frac{m_0}{2} g_{mn} \dot{x}^m \dot{x}^n - V + A_m \dot{x}^m .$$

These hold for any coordinate system; and from the appropriate one of these, and the Lagrangian equations

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{x}^n} - \frac{\partial L}{\partial x^n} = 0,$$

we obtain the relativistic or Newtonian differential equations of motion in any coordinates.

Now let us consider the Hamiltonian canonical equations.

We have already agreed to consider (A_1, A_2, A_3) as a covariant vector. We now make the same convention in regard to (π_1, π_2, π_3). Then it readily follows that the equations

$$\frac{\partial L}{\partial \dot{x}^n} = \pi_n \tag{40}$$

---

[15] Suppose that with a point P (which may be either a special point or a typical point), and with each coordinate system, we have associated an ordered triple of numbers.

If the triples of number (a_1, a_2, a_3) and $(a_{1'}, a_{2'}, a_{3'})$ associated, respectively, with any two coordinate systems (x^1, x^2, x^3) and $x^{1'}, x^{2'}, x^{3'})$ satisfy the relations

$$a_{m'} = \frac{\partial x^n}{\partial x^{m'}} a_n,$$

the numbers (a_1, a_2, a_3) are said to be the components of a covariant vector in the coordinate system (x^1, x^2, x^3). (It is understood, of course, that the partial derivatives are evaluated at the point P.)

On the other hand, if the triples of numbers (a^1, a^2, a^3) and $(a^{1'}, a^{2'}, a^{3'})$ associated with the typical coordinate systems (x^1, x^2, x^3) and $(x^{1'}, x^{2'}, x^{3'})$ satisfy the relations

$$a^{m'} = \frac{\partial x^{m'}}{\partial x^n} a^n,$$

the numbers (a^1, a^2, a^3) are said to be the components of a contravariant vector in the coordinate system (x^1, x^2, x^3).

These concepts agree only in part with the ones used in the elementary theory of vectors. From our present standpoint, the only vectors used in the elementary theories are those which are defined with reference to rectangular coordinate systems. When other coordinate systems are used (e.g. cylindrical coordinates), the vectors, defined in terms of rectangular coordinates, are merely resolved along the tangents to the coordinate curves. The components obtained in this way are not the same as the components considered in the tensor calculus, which we are using here.

are tensor equations; and since they hold when the coordinates are rectangular, they hold for all coordinate systems.[16]

We let g^{mn} denote g^{-1} times the cofactor of the element g_{mn} in the determinant

$$g = \begin{vmatrix} g_{11} & g_{12} & g_{13} \\ g_{21} & g_{22} & g_{23} \\ g_{31} & g_{32} & g_{33} \end{vmatrix}.$$

Now we write

$$H = c[m_0^2 c^2 + g^{mn}(\pi_m - A_m)(\pi_n - A_n)]^{1/2} + V \tag{41}$$

for the relativistic case, and

$$H = m_0 c^2 + (2m_0)^{-1} g^{mn}(\pi_m - A_m)(\pi_n - A_n) + V \tag{41'}$$

for the Newtonian case. We see that these expressions specialize into the ones given earlier for the Hamiltonian function when the coordinates are rectangular.

H, L, and $\pi_n \dot{x}^n$ are all scalars. Consequently, the equation

$$H + L = \pi_n \dot{x}^n \tag{42}$$

is a tensor relation; and since it holds when the coordinates are rectangular, it holds for all coordinate systems.

The Lagrangian equations can be written in the form

$$\dot{\pi}_n = \frac{\partial L}{\partial x^n}. \tag{43}$$

Let us consider the variation of the function L resulting from small variations of the x's and $\dot{x}$'s. By (40) and (43), we have the relation

$$\begin{aligned} \delta L &= \frac{\partial L}{\partial x^n} \delta x^n + \frac{\partial L}{\partial \dot{x}^n} \delta \dot{x}^n \\ &= \dot{\pi}_n \delta x^n + \pi_n \delta \dot{x}^n \\ &= \delta(\pi_n \dot{x}^n) + (\dot{\pi}_n \delta x^n - \dot{x}^n \delta \pi_n). \end{aligned} \tag{44}$$

It follows from (42) and (44) that the variation of H resulting from small variations of the x's and the π's is given by the formula

$$\delta H = \dot{x}^n \delta \pi_n - \dot{\pi}_n \delta x^n.$$

[16] The argument is explained in detail in the works on the tensor calculus cited in the bibliography.

From this it follows that we have the Hamiltonian canonical equations

$$\dot{x}^n = \frac{\partial H}{\partial \pi_n}, \qquad \dot{\pi}_n = -\frac{\partial H}{\partial x^n},$$

in any coordinate system.

We have already seen how to state Hamilton's principle in terms of general coordinates.

In the relativistic case the principle of least action takes the form: *The particle moves, in a static field of force of the type* (4), *and with the prescribed total energy E, in such a curve that the value of the integral*

$$\int_{(x^1)_1}^{(x^1)_2} \left(\left[g_{mn} \frac{dx^m}{dx^1} \frac{dx^n}{dx^1} \right]^{1/2} [(E - V)^2 c^{-2} - m_0^2 c^2]^{1/2} + A_m \frac{dx^m}{dx^1} \right) dx^1,$$

with the limits of integration held fixed, is stationary with respect to infinitesimal variations of the trajectory which leave the end points unaltered.   The corresponding form of the principle for the Newtonian case is obvious.

We are now in a position to dispose very quickly of the problem of formulating the Hamilton-Jacobi theory in terms of general curvilinear coordinates.

The general form of the Hamiltonian function being given by (41) (for the relativistic case) or (41') (for the Newtonian case), we can at once write down the partial differential equation

$$\frac{\partial W}{\partial t} + H(x^1, x^2, x^3, \partial W/\partial x^1, \partial W/\partial x^2, \partial W/\partial x^3, t) = 0. \qquad (45)$$

Let

$$W = W(x^1, x^2, x^3, t, \alpha^1, \alpha^2, \alpha^3)$$

represent the complete solution of (45), without the irrelevant additive constant of integration.

Our chief problem is that of proving that the functions $x^n(t, \alpha^1, \alpha^2, \alpha^3, \beta_1, \beta_2, \beta_3)$, $\pi_n(t, \alpha^1, \alpha^2, \alpha^3, \beta_1, \beta_2, \beta_3)$ determined by the equations

$$\frac{\partial W}{\partial \alpha^n} = \beta_n, \qquad \frac{\partial W}{\partial x^n} = \pi_n,$$

where the β's are further arbitrary constants, satisfy the canonical equations

$$\dot{x}^n = \frac{\partial H}{\partial \pi_n}, \qquad \dot{\pi}_n = -\frac{\partial H}{\partial x^n}.$$

Now, referring to the proof given in Section VIII for the special case of rectangular coordinates, we see at once that nothing in the proof depends

upon the special forms which the Hamiltonian function and equation (45) assume in those coordinates.   Hence the proof already given applies immediately to the present general case.

Similar remarks apply also to the case in which H does not involve the time explicitly, and in which we write

$$W = S - Et,$$

where S is the complete solution (without the additive arbitrary constant) of the equation

$$H(x^1, x^2, x^3, \partial S/\partial x^1, \partial S/\partial x^2, \partial S/\partial x^3) = E.$$

BIBLIOGRAPHY

Works dealing with Newtonian particle dynamics:

P. Appell, Traité de Mécanique Rationnelle, Vol. 1, 5th ed., Paris, Gauthier-Villars, 1926;

E. T. Whittaker, A Treatise on the Analytical Dynamics of Particles and Rigid Bodies, 3rd ed., Cambridge University Press, 1927.

Works dealing with the fundamental physical aspects of electron motion

N. R. Campbell, Modern Electrical Theory, 2nd ed., Cambridge University Press, 1913;

J. H. Jeans, The Mathematical Theory of Electricity and Magnetism, 5th ed., Cambridge University Press, 1925;

H. A. Lorentz, The Theory of Electrons, 2nd ed., Leipzig, B. G. Teubner, 1916;

O. W. Richardson, The Electron Theory of Matter, Cambridge University Press, 1914.

Works dealing with relativistic dynamics

A. S. Eddington, The Mathematical Theory of Relativity, 2nd ed., Cambridge University Press, 1937;

W. Pauli, Relativitätstheorie, Leipzig, B. G. Teubner, 1921;

H. Weyl, Raum-Zeit-Materie, 4th ed., Berlin, Julius Springer, 1921.

Works dealing with the Calculus of Variations

G. A. Bliss, Calculus of Variations, Chicago, Open Court Publishing Company, 1925;

E. Goursat, Cours d'Analyse Mathématique, Vol. III, 3rd ed., Paris, Gauthier-Villars, 1923.   (See also Vol. I, translated by E. R. Hedrick under the title A Course in Mathematical Analysis, Ginn and Co., 1904.)

Works on the tensor calculus

P. Appell, Traité de Mécanique Rationnelle, Vol. 5, Paris, Gauthier-Villars, 1926;

A. S. Eddington, The Mathematical Theory of Relativity, see above;

J. A. Schouten and D. J. Struik, Einführung in Die Neueren Methoden der Differentialgeometrie, Vol. I, Groningen, P. Noordhoff, 1935.

CHAPTER I

Quartz Crystal Applications

By W. P. MASON

1.1. Introduction—Purpose of Serif .

THIS paper is the first one of a series of papers dea:.·g with quartz crystals, their applications in oscillators, filters, and transducers, and the methods of producing them from the natural crystal. This series was prepared first to make available to the Western Electric Co. and other manufacturers of quartz crystals some of the specialized knowledge on these subjects that has been acquired at the Bell Telephone Laboratories. Sufficient interest has been expressed in this series to make it desirable to publish them in serial form.

This first paper in the series is a general introductory paper covering the application of crystals to oscillators, filters and transducers. An appendix is given which discusses the elastic and electric relations in crystals and gives recent measurements of the elastic constants, their temperature coefficients, and the piezoelectric constants of quartz. This paper is followed by more detailed papers by Messrs. Bond, Willard, Sykes, McSkimin, and Fair which give consideration to quartz crystallography; determination of orientation by optical methods, X-ray methods, and etching methods; the imperfections occurring in quartz crystals; modes of motion and their calculation; the dimensioning of crystals to avoid undesirable resonances; and the use of crystals in oscillators.

1.2 Early History of Piezoelectricity and its Applications

The direct piezoelectric effect was discovered by the brothers Curie in 1880. They measured the effect first for a quartz crystal by putting a weight on the surface and measuring the charge appearing on the surface, the magnitude of which was proportional to the applied weight. A simple model for demonstrating this effect can be made by using a large piece of Rochelle salt cut with its length $45°$ from the Y and Z crystallographic axes and placing tinfoil electrodes normal to the X axis. If these electrodes are connected to a neon lamp, and the crystal is compressed by hitting it with a hammer, a charge is generated on the surface and a voltage applied to the lamp sufficient to break it down. In fact as much as 2000 volts can be generated by striking the crystal hard.

The converse piezoelectric effect was predicted in 1881 by the French physicist Lippmann on the basis of the principle of conservation of electricity. It was verified in the same year by the brothers Curie. In this effect a crystal is strained when a voltage is applied to it. The effect can be demonstrated by a model which consists of two thin pieces of Rochelle salt poled so that one expands when the voltage is applied and the other contracts. The result is—as in a bimetallic thermostat—the crystal bends. For crystals 10 mil inches thick and 4 inches long, a ninety-volt battery applied causes a displacement of a quarter of an inch or more of the end of the unit. Reversing the voltage reverses the direction of the displacement. The Curies constructed a bimorph unit of this type out of quartz and used it practically to measure voltage by measuring the displacement of the end of the crystal. By connecting the leads of an electrometer to the terminals, they could measure force applied by measuring the amount of charge generated at the terminals.

Outside of this use which was quite minor, the piezoelectric effect remained a scientific curiosity until the war of 1914–1918. It did inspire, however, considerable scientific speculation. Lord Kelvin in 1893 proposed a model for explaining the piezoelectricity of quartz and was able to calculate approximately the value of the piezoelectric constant. This model is discussed briefly in the next section. He also constructed and demonstrated a "piezoelectric pile" made from small spheres of zinc and copper, to illustrate the effect. At about the same time (1890–1892) Voigt published a series of papers followed by a book "Lehrbuch der Kristall Physik" (1910) in which the stresses, strains, fields and polarizations of piezoelectric crystals are related in mathematical form. These mathematical expressions (which are discussed further in the appendix) form a basis for the development of the properties of oriented crystals as discussed in section 1.5.

During the war of 1914–1918, Professor Langevin in Paris was requested by the French Government to devise some way of detecting submarines by acoustic waves they produce in water. After trying several devices he finally found that piezoelectric quartz plates could be used for that purpose. His device, which is shown in Fig. 1.1, consisted essentially of a mosaic of quartz which has the property that when a voltage is applied the crystal will expand and send out a longitudinal wave. Similarly, if a wave strikes it, the wave will set the quartz in vibration and generate a voltage which can be detected by vacuum tube devices. Langevin did not get his device perfected till after the war so it was not used at that time to detect submarines. Similar devices have, however, been used in this war. Langevin's original apparatus was used extensively as a sonic depth finder. In this use a pulse is generated which is recorded directly on a moving record and

is also sent out into the ocean. It strikes the bottom and is reflected back causing another mark to appear on the record. Knowing the difference in time and the velocity of sound in sea water, the distance to the bottom can be measured. A typical record is shown in Fig. 1.2. The top record shows the contour of the sea bottom while the second record shows the reflections from a school of fish.

At about the same time, Nicolson at Bell Telephone Laboratories was experimenting with Rochelle salt, another piezoelectric material having a
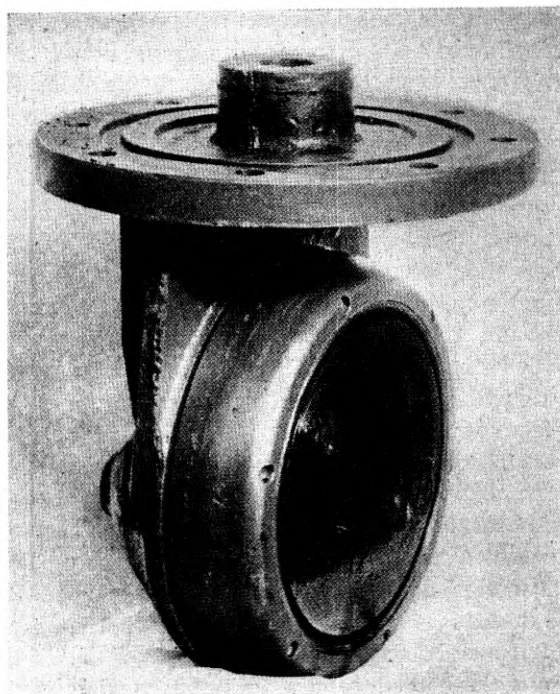
Fig. 1.1—Ultrasonic transmitting apparatus

much larger piezoelectric effect than quartz. He constructed and demonstrated loud speakers, microphones, and phonograph pick-ups using Rochelle salt.[1] He was also the first one to control an oscillator by means of a crystal —in this case Rochelle salt—and has the primary crystal oscillator patent.[2] Nicolson's circuit is shown in Fig. 1.3. The crystal is effectively in a path between the resonating coil in the output and the grid, since the electrode

[1] "The Piezoelectric Effect in the Composite Rochelle Salt Crystal"—A. M. Nicolson, *Proc. A. I. E. E.* 1919, 38, 1315.
[2] See Patent 2,212,845 filed April 10, 1918; issued Aug. 27, 1940.

3 is in the direction of the smallest piezoelectric effect in Rochelle salt and contributes little to the action. If terminal one to the tapped coil is at the top of the coil, the circuit although employing a three electrode crystal connection, effectively reduces to *B* in which the crystal is in the feedback path from plate to grid. On the other hand, if the tap is effectively at the bottom of the coil, the crystal is between grid and ground and feedback occurs through a distributed capacity from plate to grid. Both of these
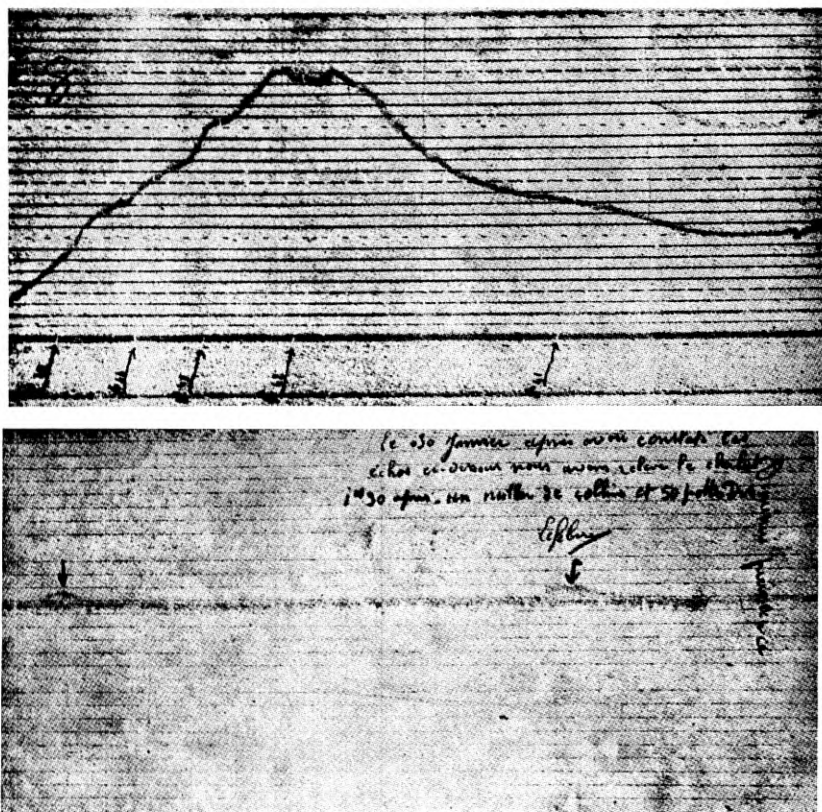
Fig. 1.2—Ocean contour curves

circuits *B* and *C* are widely used in oscillators of Pierce. Prof. G. W. Pierce published a circuit similar to circuit *B*, having a two electrode quartz crystal connected between grid and plate.[3]

In 1921, Professor Cady at Wesleyan University first showed[4] that quartz

[3] "Piezoelectric Crystal Resonators and Crystal Oscillators Applied to the Precision Calibration of Wave Meters," G. W. Pierce, Amer. Acad. of Arts and Sciences, Oct. 1923, 81–106.
[4] "The Piezoelectric Resonator" W. G. Cady, *Proc. I. R. E.* 1922, 10 83.

crystals could be used to control oscillators and that much more stable oscillators could be obtained in this fashion. These were later applied to controlling the frequency of broadcasting stations and radio transmitters in general and about 1925 Mr. W. A. Marrison applied them to obtain a very constant frequency and time standard, which is now used considerably by the Bell System, by radio broadcasting systems, and by power companies. The oscillators were subsequently improved by using crystals with small temperature coefficients as described in Section V. At the present time
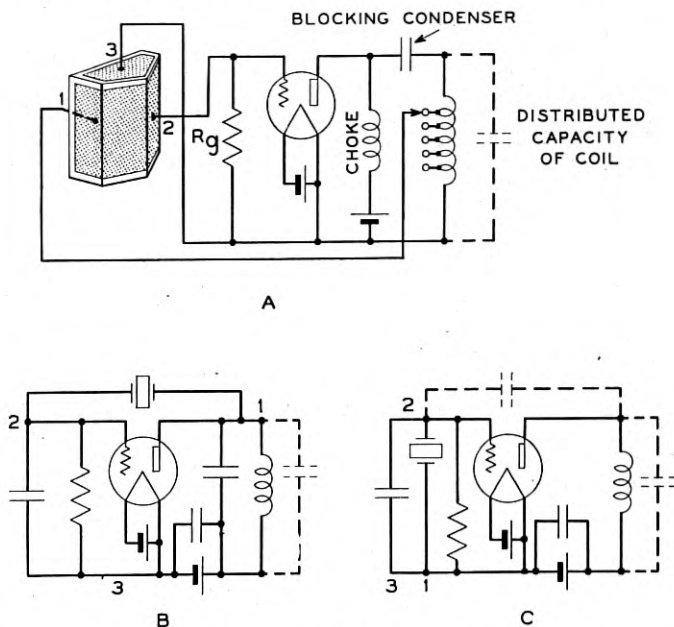
Fig. 1.3—Nicolson's oscillator circuit

crystal controlled oscillators are used very widely in radio military and commercial applications.

Another large use for quartz crystals is their use in providing very selective filters. Probably the first use of a crystal to select a narrow frequency range was made by Cady.[4] Using the very sharp maximum in current through a crystal at its resonant frequency, Cady proposed the use of such a crystal as a wave standard. This is equivalent to the use of a crystal as a tuned circuit. By incorporating a crystal in a three-winding transformer and balancing out the static capacity of the crystal by an auxiliary condenser, W. A. Marrison[5] improved the selecting ability of a crystal used

[5] Patent 1,994,658, filed June 7, 1927; issued March 19, 1935.

as a narrow band filter. At about the same time, L. Espenschied,[6] taking advantage of the knowledge of the equivalent electrical circuit of a crystal given previously by Van Dyke,[7] showed how to combine other electrical elements with crystals in ladder form to obtain band-pass filters. It was not, however, until the crystals were combined with auxiliary coils and condensers into the form of resistance compensated lattice type networks[8] that much progress was made in achieving the wide pass-band characteristics necessary for telephone and radio communication. Such filters have provided very selective devices which are able to separate one band of speech frequencies from another band different by only a small frequency percentage from the desired band. This property makes it possible to space channels close together with only a small frequency separation up to a high frequency, and such filters have had a wide use in the high-frequency carrier systems, and in the coaxial system which transmits more than 480 conversions over one pair of conductors. In radio systems such filters have been used extensively in separating one side band from the other in single sideband systems.

In conclusion we can say, that the science of piezoelectricity was born about 62 years ago, lay dormant for nearly 40 years, but during the last 25 years has advanced at such a rate that it can be regarded as one of the foundation stones of the whole communication art.

1.3. THEORY OF PIEZOELECTRIC MATERIALS

Piezoelectric crystals are of interest in communication circuits because they possess three properties. These properties are: (1) the piezoelectric effect provides a coupling between the electrical circuit and the mechanical properties of the crystal; (2) the internal dissipation of most crystals and particularly quartz crystals is very low, and the density and elastic constants of the crystals are very uniform, so that a crystal cut at a given orientation always has the same frequency constant; and (3), at specified orientations crystals can be cut which have advantageous mechanical properties such as a small change in frequency with a change in temperature, or a freedom from secondary modes of motion. It is the purpose of this section to discuss the first property, the coupling between the electrical and mechanical properties of the crystal.

The piezoelectricity of quartz and other materials is due to the fact that

---

[6] Patent 1,795,204, filed Jan. 3, 1927, issued August 8, 1933.

[7] K. S. Van Dyke; Abstract 52, *Phys. Rev.* June 1925; *Proc. I. R. E.* June 1928.

[8] See "Electrical Wave Filters Employing Quartz Crystals as Elements," W. P. Mason, *B. S. T. J.*, Vol. XIII, p. 405, July 1934; "Resistance Compensated Band Pass Crystal Filters for Unbalanced Circuits," *B. S. T. J.*, Vol. XVI, p. 423, Oct. 1937; "The Evolution of the Crystal Wave Filter," O. E. Buckley, *Jour. App. Phys.*, Oct. 1936; and Patents 1,921,035; 1,967,249; 1,967,250; 1,969,571; 1,974,081; 2,045,991; 2,094,044.

a pressure which deforms the crystal lattice causes a separation of the centers of gravity of the positive and negative charges thus generating a dipole moment (product of the value of the charges by their separation) in each molecule. How this separation can cause a coupling to an electrical circuit is illustrated by Fig. 1.4 which shows a crystal with metal electrodes normal to the direction of charge separation. If we short-circuit these electrodes and apply a stress which causes the centers of gravity of the charges to separate, free negative charges in the wire will be drawn toward the electrode in the direction of positive charge separation, and free positive charges in the wire will be drawn to the electrode in the direction of negative charge displacement until the crystal appears to be electrically neutral by any test conducted outside the crystal. When the stress is released the charges in the wire will flow back to their normal position. If, during the process, we connect an oscillograph in the short-circuited wire, there will be a pulse of current in one direction when the stress is applied and a pulse in the oppo-
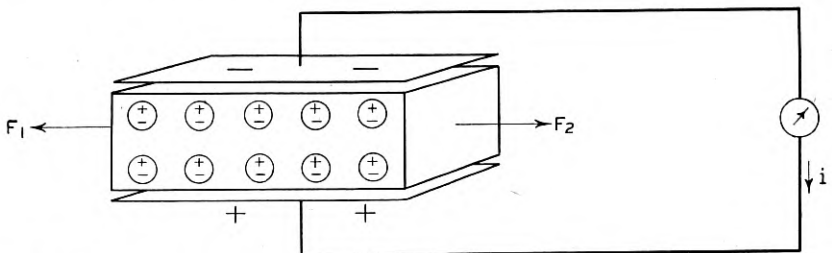
Fig. 1.4—Method for transforming mechanical energy into electrical energy in a crystal

site direction when the stress is released. By putting a resistance in the connecting wire and applying a sinusoidal stress to the crystal, an alternating current will flow through the load and consequently mechanical power will be changed into electrical power. Using the converse effect, a source of alternating voltage in the electrical circuit will produce an alternating stress in the crystal, and if this is working against a mechanical load, the electrical energy will be changed into mechanical energy.

To apply this concept to quartz let us consider Fig. 1.5, which represents the approximate arrangement of molecules in a quartz molecule. Lord Kelvin's explanation of the piezoelectricity of quartz is the following:

"The diagram (Fig. 1.5A) shows a crystalline molecule surrounded by six nearest neighbors in a plane perpendicular to the optic axis of a quartz crystal. Each silicon atom is represented by + (plus) and each oxygen double atom — (minus). The constituents of each cluster must be supposed to be held together in stable equilibrium in virture of their chemical affinities. The different clusters, or crystalline molecules, must be supposed to be relatively mobile before taking

positions in the formation of a crystal. But we must suppose, or we may suppose, the mutual forces of attraction (or chemical affinity), between the silicon of one crystalline molecule and the oxygen of a neighboring crystalline molecule, to be influential in determining the orientation of each crystalline molecule, and in causing disturbance in the relative positions of the atoms of each molecule, when the crystal is strained by force applied from without.

"Imagine now each double atom of oxygen to be a small negatively electrified particle, and each atom of silicon to be a particle electrified with an equal quantity of positive electricity. Suppose now such pressures, positive and negative, to be applied to the surface of a portion of crystal as shall produce a simple elongation in the direction perpendicular to one of the three sets of rows. This strain is indicated by the arrow heads in Fig. 1.5A and is realized to an exaggerated extent in Fig. 1.5B.

(A) (B)
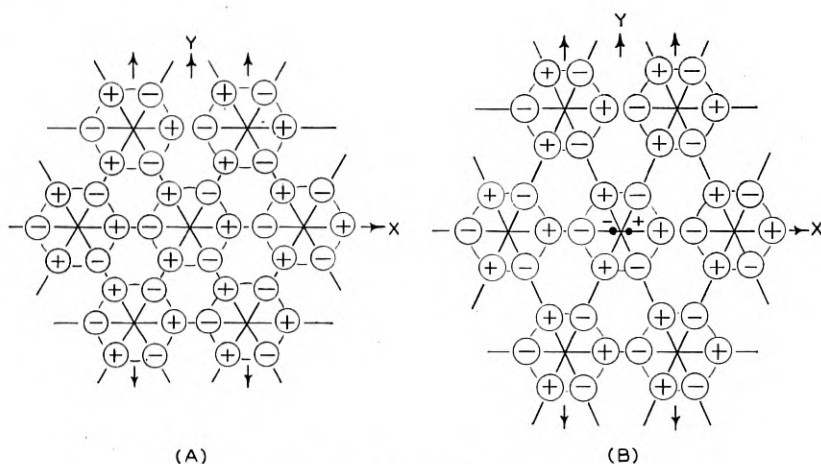
Fig. 1.5—Kelvin's model of quartz molecules

"This second diagram shows all the atoms and the centers of all the crystalline molecules in the positions to which they are brought by the strain. Both diagrams are drawn on the supposition that the stiffness of the relative configuration of atoms of each molecule is slight enough to allow the mutual attractions between the positive atoms and the negative atoms of neighboring molecules to keep them in line through the centers of the molecules, as Fig. 1.5A shows for the undisturbed condition of the systems, and Fig. 1.5B for the system subjected to the supposed elongation. Hence two of the three diameters through atoms of each crystalline molecule are altered in direction, by the elongation, while the diameter through the third pair of atoms remains unchanged, as is clearly shown by Fig. 1.5B compared to Fig. 1.5A.

"Remark, first that the rows of atoms, in lines through the centers of the crystalline molecules, perpendicular to the direction of the strain, are shifted to parallel positions with distances between the atoms in them unchanged. Hence the atoms in these rows contributed nothing to the electrical effect. But in parallel to these rows, on each side of the center of each molecule, we find two pairs of atoms whose distances are diminished.

"This produces an electrical effect which, for great distances from the molecule, is calculated by the same formula as the magnetic effect of an infinitesimal bar magnet whose magnetic moment is numerically equal to the product of the quantity of electricity of a single atom into the sum of the diminutions of the two distances between the atoms of the two pairs under consideration. Hence, denoting by N the number of crystalline molecules per unit bulk of the crystal; by b the radius of the circle of each crystalline molecule; by q the quantity of electricity of each of the six atoms or double atoms, whether positive or negative; by θ the change of direction of each of the two diameters through atoms which experience change of direction; and by μ the electric moment developed per unit volume of the crystal, by the strain which we have been considering and which is shown in Fig. 1.5B; we have

$$\mu = N q \, 4b \, \theta \cos 30° = 2\sqrt{3} \, N \, b \, q \, \theta \qquad (1.1)"$$

Kelvin's model shows some of the symmetry properties of quartz. The axis marked X is the X or electrical axis of the crystal. The Z or optic axis is normal to the plane of the paper. The Y or mechanical axis is the axis

(A)   (B)
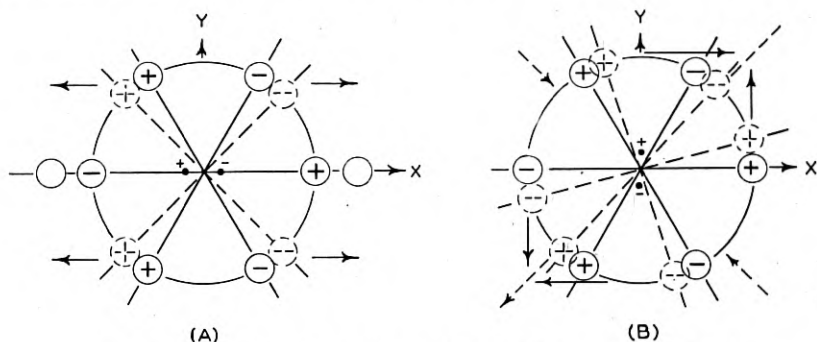
Fig. 1.6—Longitudinal and shear strains applied to a quartz molecule

along which the stress is applied. It is obvious that if we rotate the direction of the applied stress by 120°, a similar separation of charges at right angles to the stress will occur. There are then three electrical axes and three mechanical axes so that the optic axis can be regarded as an axis of threefold symmetry for the crystal.

As can be shown from an extension of Kelvin's model there are two other types of stresses that will produce a charge separation normal to the axis. Suppose that we stress the crystal along the X or electrical axis as shown by Fig. 1.6A. Applying the same reasoning as before, we see that the apex molecules are separated farther apart without changing the separation between the other molecules. This results in a separation of the centers of gravity of the positive and negative charges, with the negative charges moving toward the left and the positive charges moving toward the right. The separation is still along the electric axis, but is in the opposite direction

to that caused by a stress along the Y axes. A detailed analysis shows that the value of the electrical separation moment (dipole moment) for a stress along either axis is the same value but the sign is reversed. A longitudinal stress then can only produce a charge moment along the X or electrical axis which is the origin of the name electrical axis.

If, however, we introduce a different kind of stress known as a shearing stress, a separation of centers of charges can occur along the mechanical or Y axis of the crystal. A simple shear stress is one in which forces act normal to the direction of space separation rather than along it as shown, for example, by the two opposed arrows normal to the mechanical axis in Fig. 1.6B. Such a shear does not occur in nature, but rather a pure shear which consists of two simple shears which are directed in such a way as to produce no rotation of the molecule as a whole about its axis. If we resolve these force components along directions 45° from the crystal axes, a pure shear is equivalent to an extensional stress along one 45° axis and a compressional stress along the other 45° axis. Such a stress would cause the charges to be displaced from their normal position, as shown in the figure. This causes the center of positive charge to be displaced downward along the mechanical or Y axis of the crystal while the center of negative charge is displaced upward along the mechanical axis.

These three relations can be written in the form

$$P_x = -d_{11}X_x + d_{11}Y_y ; \qquad P_y = 2d_{11}X_y \qquad (1.2)$$

where P_x is the polarization or charge per unit area developed on an electrode surface normal to the X axis due to the applied longitudinal stresses X_x and Y_y, while P_y is the polarization normal to the Y axis caused by the shearing stress X_y. d_{11} is the piezoelectric constant and equations (1.2) show that the magnitudes of all these effects are closely related. In addition to these three major piezoelectric effects, quartz has two smaller effects which, since they are connected with the distribution of molecules in the YZ and XZ planes, cannot be demonstrated by the figures given previously. The complete piezoelectric relations are then

$$P_x = -d_{11}X_x + d_{11}Y_y - d_{14}Y_z ; \qquad P_y = d_{14}Z_x + 2d_{11}X_y \qquad (1.3)$$

where Y_z and Z_x are respectively similar shearing stresses exerted in the YZ and ZX planes respectively. The best values for the d_{11} and d_{14} constants are respectively

$$d_{11} = -6.76 \times 10^{-8} \frac{\text{e.s.u.}}{\text{dyne}}; \qquad d_{14} = 2.56 \times 10^{-8} \frac{\text{e.s.u.}}{\text{dyne}} \qquad (1.4)$$

as obtained by recent measurements for a number of X cut and rotated X-cut crystals discussed in appendix A.

Quartz is not the only type of crystal which is piezoelectric. In fact there are hundreds of crystals that exhibit this property. Whether a crystal is piezoelectric or not and the relation between the stresses and charge displacements depend on the symmetry of the crystal. Whenever there is a center of symmetry; that is, when the properties of the crystal are the same in both directions along any line, no piezoelectric effect can occur. This is illustrated by the simple arrangement of atoms shown by Fig. 1.7. It is obvious that no symmetrical application of forces can separate the center of gravity of the charges and hence such a crystal will not be
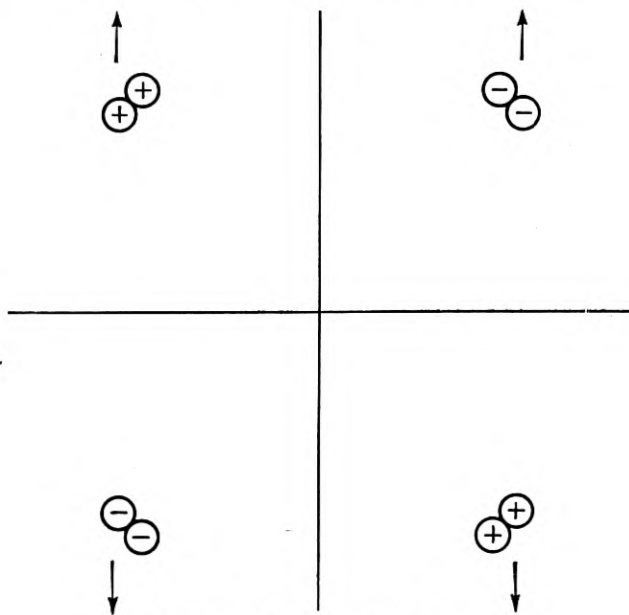
Fig. 1.7—Crystal with a center of symmetry

piezoelectric. Crystals can be classified into 32 possible classes on the basis of the symmetry they exhibit; and of these 32 classes, 20 are piezoelectric and 12 are not. As illustrated by the model for quartz, the response to different types of force depends solely on the type of symmetry existing in the crystal.

1.4. ELECTRICAL IMPEDANCE AND LOW DISSIPATION IN CRYSTALS

The first crystal used by Cady in controlling oscillators, was a crystal cut with its major faces perpendicular to the X or electrical axis and with its length along the Y or mechanical axis. Referring to Fig. 1.5B, we see that a stretch along the Y axis will produce a charge displacement along the E

or X axis. Conversely, a voltage applied along the X axis will produce a charge displacement and consequently a mechanical stress along the Y axis which will set up a longitudinal wave along the mechanical axis. As shown by Fig. 1.8, the type of motion resulting when the crystal is free to move on the ends is one in which the center is stationary and the ends move in and out. The crystal can then be clamped at its center or mounted from leads soldered to electrodes deposited on the surface.

In using a crystal in an electrical circuit it is desirable to have an electrical equivalent circuit which will represent the electrical impedance as measured from the terminals of the crystal. Such a circuit[9] is shown in Fig. 1.8. In

$$f_R = \frac{1}{2l}\sqrt{\frac{1}{\rho S_{22}}} = \frac{V}{2l} \; ; \; S_{22} = \frac{1}{Y}$$

f_A GIVEN BY SOLVING

$$\frac{\pi f_A l}{V} \cot \frac{\pi f_A l}{V} = -(d_{12})^2 \frac{4\pi}{K S_{22}} = -k^2$$
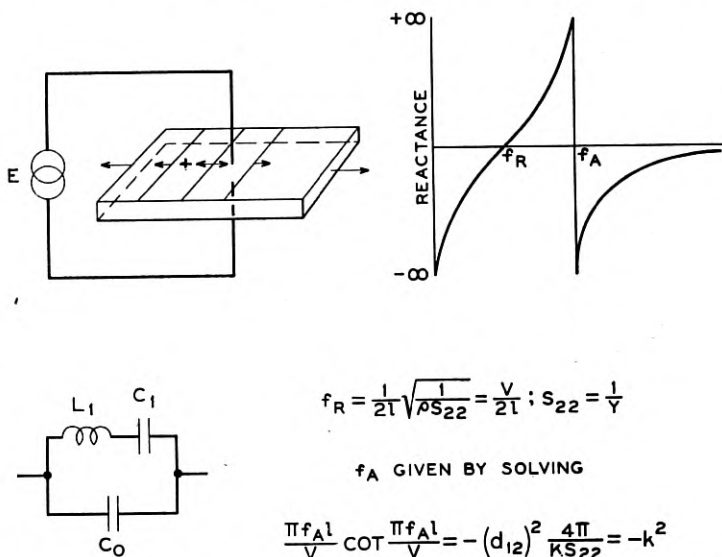
Fig. 1.8—Longitudinally vibrating crystal and electrical equivalent circuit

this representation C_0 is the static capacity of the crystal which would be measured if the crystal were held from moving. C_1 is the stiffness of the crystal transformed into electrical terms through the piezoelectric effect of the crystal, while L_1 is the effective mass of the crystal also transformed into electrical terms. The resonant frequency of the crystal is determined by the Young's modulus and density of the bar according to the usual formula:

$$f_R = \frac{1}{2\ell}\sqrt{\frac{Y_0}{\rho}} \tag{1.5}$$

[9] Circuits of this type for representing the electrical impedance of a crystal were first derived by Van Dyke; see reference (7). The method of deriving them from Voigt's equations is discussed in the appendix.

where Y_0 is the value of Young's modulus along the bar, ρ the density, and ℓ the length of the bar.

A significant feature of the equivalent circuit is that there is always a definite ratio between C_0 and C_1 for a given crystal cut. This is really a measure of the ratio of electrical to mechanical energy stored in the crystal under an applied constant voltage. The reactance characteristic of the network is shown by Fig. 1.8 as a function of frequency. The reactance starts out as a negative reactance at low frequencies, becomes zero at the resonant frequency f_R, becomes positive and very large at the anti-resonant frequency f_A, then again becomes a negative reactance. Due to the high ratio of C_0 to C_1 existing in a crystal the separation between f_A and f_R becomes very small. For example, for an AT crystal this ratio is around 200 and the separation of f_A from f_R is only a quarter of a per cent in frequency. Since it can be shown that an oscillator will only oscillate on the positive reactance part of the crystal characteristic, the narrow separation between resonant and anti-resonant frequencies explains why a crystal can act as such a good stabilizer for an oscillator. As long as the crystal resonance itself does not change with temperature or other conditions, the very sharp reactance frequency characteristic will not allow the oscillator frequency to change much with a change in oscillator voltage, tube conditions, or any other changes which are likely to cause a change in frequency for a coil and condenser controlled oscillator.

Strictly speaking, a resistance should be added in series with the inductance L_1 to represent the internal losses in the crystal, the loss of energy at the clamping points and the loss of energy due to setting up of air waves by the crystal motion. However, the value of this resistance and the amount of energy lost is very small in a crystal compared to what the losses are in purely electrical elements. A demonstration which shows this effect and shows that most of the losses of a well mounted longitudinally vibrating crystal are acoustic losses caused by setting up air waves in the vicinity of the crystal, can be made by using two oscillators, one a fixed oscillator and the other one controlled by a resonant circuit or a crystal. The fixed oscillator may be set at 99 kilocycles and the crystal oscillator controlled by a 100-kc crystal. The two will beat together giving the 1000-cycle note. When the battery is taken off the crystal oscillator, it continues to oscillate till the energy built up in the crystal is dissipated in the internal dissipation of the crystal. A good electrical circuit which has a ratio of reactance to resistance, or Q of the coil of 300 dies down almost instantaneously. For a crystal mounted in air it takes about half a second to become inaudible. This corresponds to a Q of 30,000 where Q is defined as the ratio of the reactance of the coil L_1 of Fig. 1.8 to the resistance. For a crystal mounted in a vacuum a much higher Q is obtained due to the elimination of the loss of energy by acoustic radiation. For such a crystal it takes eight

seconds to die down which corresponds to a Q of 330,000, which is about 1000 times as great as that for a good electrical circuit.

1.5. MODES OF MOTION AND CRYSTAL ORIENTATION TO PRODUCE LOW TEMPERATURE COEFFICIENT CRYSTALS

As mentioned previously the first crystal cut used in oscillators was a longitudinal vibration along the Y or mechanical axis excited by a field applied
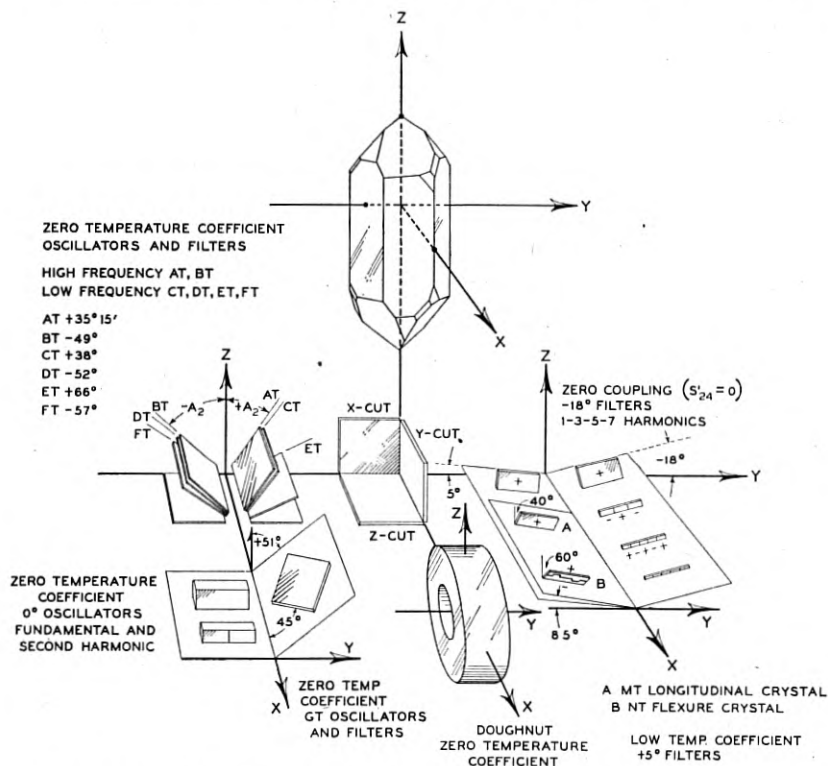

ZERO TEMPERATURE COEFFICIENT
OSCILLATORS AND FILTERS

HIGH FREQUENCY AT, BT
LOW FREQUENCY CT, DT, ET, FT

AT +35° 15′
BT −49°
CT +38°
DT −52°
ET +66°
FT −57°

ZERO TEMPERATURE COEFFICIENT
0° OSCILLATORS
FUNDAMENTAL AND
SECOND HARMONIC

X-CUT
Y-CUT
Z-CUT

ZERO COUPLING $(s'_{24} = 0)$
−18° FILTERS
1-3-5-7 HARMONICS

ZERO TEMP
COEFFICIENT
GT OSCILLATORS
AND FILTERS

DOUGHNUT
ZERO TEMPERATURE
COEFFICIENT

A MT LONGITUDINAL CRYSTAL
B NT FLEXURE CRYSTAL

LOW TEMP. COEFFICIENT
+5° FILTERS

Fig. 1.9—Oriented quartz crystal cuts in relation to the natural crystal

along the electrical or X axis. This mode gives a good resonance free from other modes and a modification of it is now used in crystal filters. This modification, as shown by Fig. 1.9, ($-18.5°$ filter crystal) consists in rotating the direction of the length by 18.5° from the Y or mechanical axis, about the X or electrical axis. As described previously[10], the effect of this rotation is to eliminate the coupling between the desired longitudinal mode and the undesired face shear mode, thus simplifying the motion and eliminating an

[10] "Electrical Wave Filters Employing Quartz Crystals as Elements," W. P. Mason, *B. S. T. J.*, Vol. XIII, p. 405 July 1934 or patent 2,173,589.

undesired resonance. However, to get a reasonably high frequency out of a length vibrating type of cut requires too small a length to be practical.

It was not long before crystal oscillators were controlled by thickness vibrating crystals whose frequencies were determined by the thickness of the crystals or by their smallest dimension. Referring to Fig. 1.6A, we see that the same X cut type of crystal will generate a vibration along the electrical or X axis when a field is applied along this axis. Since the thickness dimension can be made very small, a high frequency is obtainable. However, when the smallest dimension is used to control the frequency, a difficulty arises not present when the largest dimension is used to control the frequency, namely, that harmonics and overtone modes of all the lower frequency types of motion produce frequencies near the frequency of the thickness mode and it is difficult to pick out the desired mode. This was
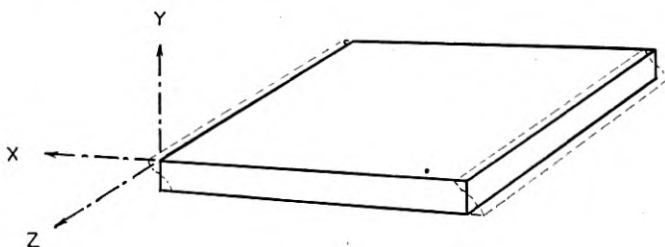
$$\delta X = A \cos \frac{n \pi y}{a} \cos 2n\pi f t$$

$$f = \frac{n}{2t} \sqrt{\frac{c_{66}}{\rho}} \; ; \quad n = 1, 3, 5 \; ; \qquad c_{66} = \frac{c_{11} - c_{12}}{2}$$

Fig. 1.10—High frequency shear mode of motion

especially true for the thickness vibrating X cut crystal and led to its abandonment in favor of Y cut crystals vibrating in shear.

As seen from Fig. 1.6B, when a voltage is applied along the Y or mechanical axis, a shear vibration is produced which tends to change a square into a rhombus. For a large plate in which the edge dimensions are large compared to the thickness, the motion occurs as shown by Fig. 1.10. For such a plate the motion is perpendicular to the thickness, which is the direction of transmission of the wave, and hence a shear wave is sometimes called a transverse wave. The frequency of such a wave can be shown to be

$$f = \frac{1}{2t} \sqrt{\frac{c_{66}}{\rho}} \tag{1.6}$$

where t is the thickness of the plate, c_{66} is the shear stiffness constant and ρ the density. The use of Y cut plates considerably improved the per-

formance of oscillators since the plates do not have as many secondary modes of motion as do the X plates. They have, however, one drawback. The frequency increases about 86 parts in a million for every degree Centigrade increase in temperature. This requires regulating the temperature quite closely.

In order to improve on the performance of the Y cut crystal, investigations were made by Lack, Willard and Fair, Koga, Bechmann, Straubel and others[11] on how the properties of such crystals varied as the orientation angle of cutting blanks from the natural crystal was varied. As shown by Fig. 1.9, the crystals investigated all had one edge along the X or electrical axis with the normals making positive and negative angles with the Y axis. All of these crystals will have a component of field along the Y axis, which
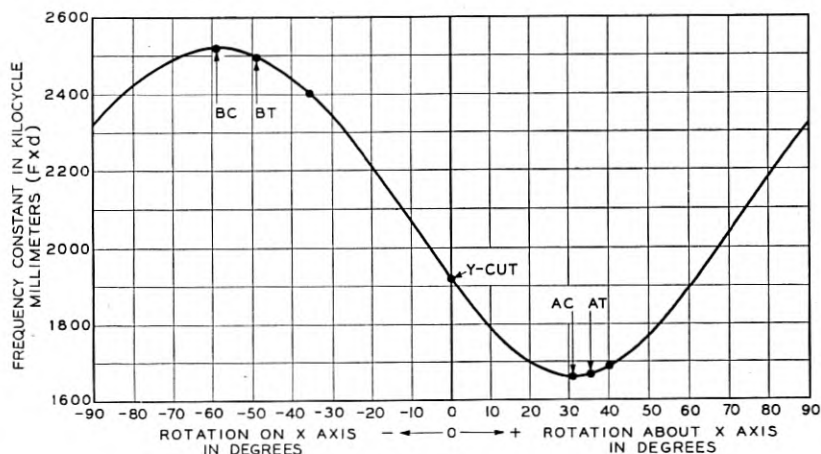
Fig. 1.11—Frequency constant of oriented V cut crystals

will produce a shearing motion until the angles of cut approach 90 degrees from the Y axis. The smaller the angle A_2 the more strongly will the shear mode be driven. However, advantageous elastic relationships can be obtained by using oriented cuts. As shown by Fig. 1.11, Lack, Willard and Fair found that the frequency constant of a rotated crystal expressed in kilocycle millimeters varied with angle of cut and that there was a minimum frequency at +31 degrees and a maximum at −59 degrees. It was subsequently pointed[12] out that these minimum and maximum points were significant angles in the elastic behavior of the crystal for they were the angles for

---

[11] "Some Improvements in Quartz Crystal Circuit Elements," F. R. Lack, G. W. Willard, I. E. Fair—*B. S. T. J.*, Vol. 13, pp. 453–463, July 1934; R. Bechmann—*HF Techn. u. El. Ak.* 44, 145 (1934); I. Koga—*Rep. of Rad. Res. i. Jap.* 6, 1 (1934); J. Straubel, *Z. tech. Physik.*, 35, 179, 1934.
[12] See patent 2,173,589.

which the high-frequency shear mode had zero coupling with the trouble-some low-frequency shear mode system of vibrations. Crystals cut at these angles have a much cleaner frequency spectrum than Y cut crystals. Lack, Willard, and Fair also found that the temperature coefficient of frequency varied with angle as shown by Fig. 1.12. Starting from a high positive value for the Y cut, the coefficient becomes zero at an angle of $+35° - 15'$ and $-49°$. The first angle cut is known as the AT cut and the second as the BT cut. Since the AT angle is nearer the Y cut, the piezo-electric constant is larger and it is more strongly driven than the BT. On the other
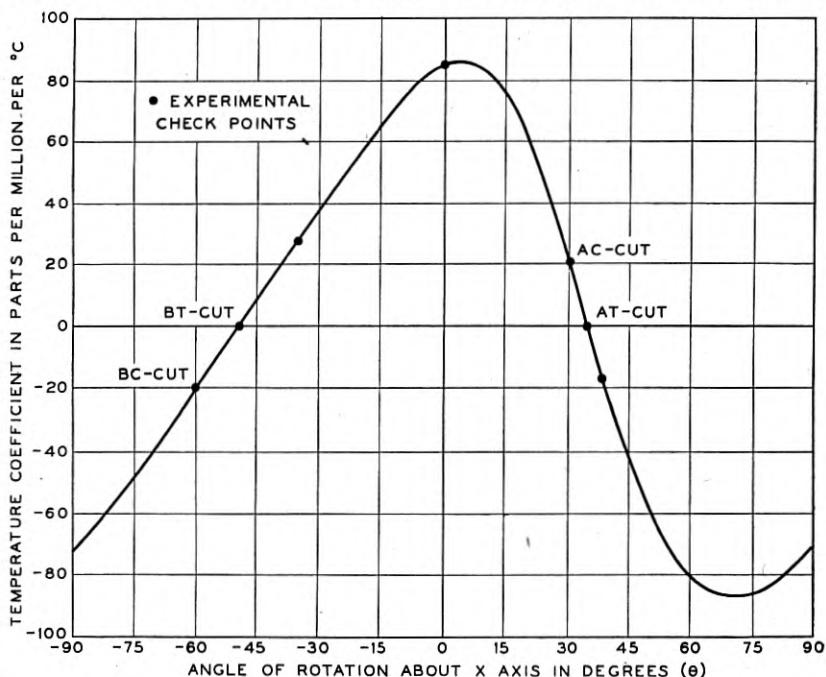
Fig. 1.12—Temperature coefficients of oriented Y cut crystals

hand, the BT has a higher frequency for the same thickness. Both crystals are near enough to the AC and BC cuts so that the systems of low-frequency shear modes are rather weakly driven. On the other hand, the shear mode of both crystals is rather strongly coupled to flexure modes of motion, as will be discussed by Mr. Sykes in a later chapter, and the crystal has to be exactly dimensioned in order that the flexure frequencies and other disturbing frequencies will not coincide with the desired shear mode.

Other oriented shear crystals for lower frequency work are the CT and DT crystals investigated by Willard and Hight. They are related to the AT and BT crystals as shown by Fig. 1.13. The plate on the right shows the

motion of an *AT* plate.   If we were to increase the thickness dimension until the plate was nearly square, the *AT* motion would correspond to a face shear mode which should be controlled by the same elastic constants as the *AT* motion.   At the same time in order to drive the crystal efficiently we could decrease the width until it became the thickness.   This procedure would be the same as cutting a crystal at right angles to the *AT* and would suggest that by so doing we should obtain a low-frequency shear crystal with a low coefficient.   Actually, Willard and Hight found that a crystal
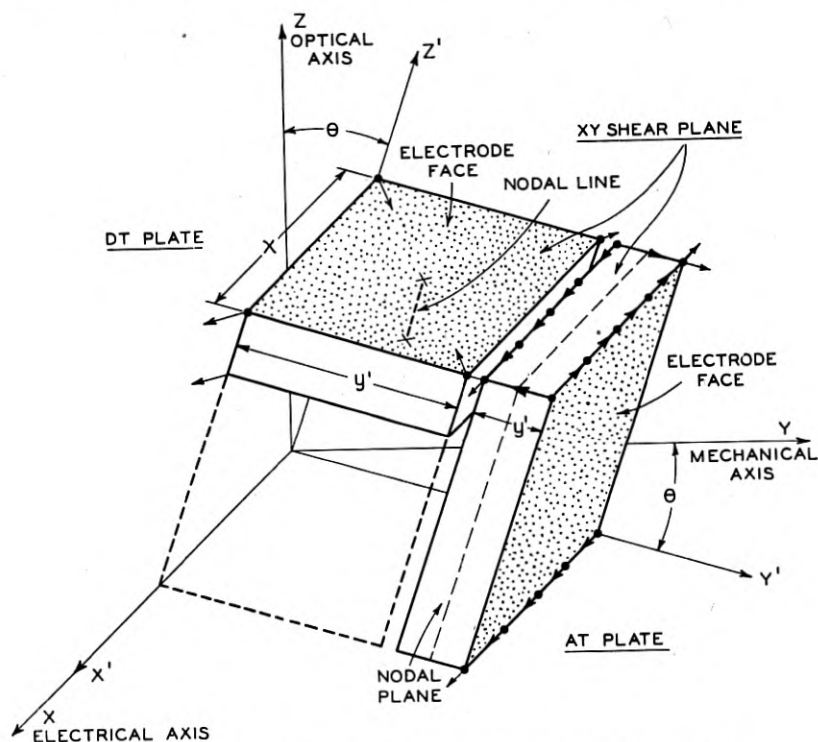
Fig. 1.13—Relation of *AT* and *DT* low temperature coefficient crystals

at −52° or 87° from the *AT* would give a low coefficient.   This crystal was called the *DT*.   Similarly, a crystal cut at +38° or 87° from the *BT* would also give a low coefficient and this has been called the *CT*.   It can be shown that a component of the voltage applied along the mechanical axis will drive the shearing type of motion.   The *CT* is larger for the same frequency and more strongly driven than the *DT*.   It is extensively used in controlling oscillators in the frequency range from 200 to 500 kilocycles.

Quite a few other crystal cuts have been discovered as shown by Fig. 1.9.

Another important cut is the GT,[13] which has a very constant frequency over a wide temperature range. As shown by Fig. 1.14, all zero temperature coefficient crystals are zero coefficient at one temperature only and usually vary in a square law curve about this temperature. The GT crystal represented an attempt to get a crystal in which the frequency remained constant over a wide temperature range. As can be seen from the figure, when properly adjusted this aim is attained, for the frequency does not vary more than one part in a million over a 100-degree Centigrade range of temperature.

This crystal makes use of the fact that a face shear vibration can be resolved into two longitudinal vibrations coupled together. As shown by
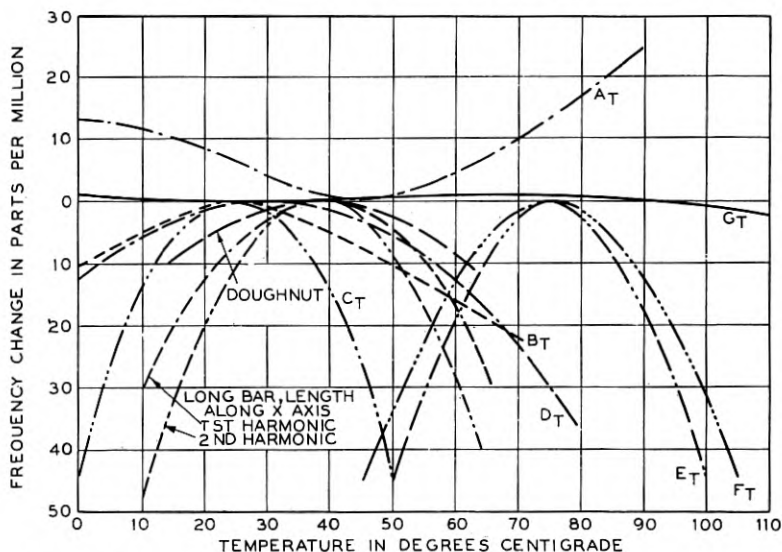
Fig. 1.14—Temperature frequency characteristics of a number of low temperature coefficient crystals

Fig. 1.15, if we cut a crystal at an angle of 45 degrees from that of a shear vibrating crystal, an expansion occurs along one axis and a contraction along the other indicating that a face shear can be resolved into two longitudinal modes that are coupled together. Now since it can be shown that all pure longitudinal modes for blanks cut in all possible directions in a quartz crystal will have zero or negative temperature coefficients,[14] it follows that if we have a shear vibrating crystal with a positive coefficient, that

[13] "A New Quartz Crystal Plate, Designated the GT, Which Produces a Very Constant Frequency Over A Wide Temperature Range," W. P. Mason, *Proc. I. R. E.*, Vol., 28 pr. 220–223, May 1940

[14] This can be proved as discussed in the appendix by combining the Voigt expressions for the elastic relations in a crystal with the measured temperature coefficients of the six elastic constants.

coefficient must have been caused by the coupling between the two modes. As a result of this observation it follows that if we have a shear vibrating crystal with a positive temperature coefficient and cut another crystal at 45 degrees from this crystal, the strong coupled mode which corresponds to the shear vibration will also have a positive temperature coefficient. As we grind down on the side, the two modes become farther apart in frequency and less closely coupled. Then, since they both will have a negative coefficient if separated far enough, it follows that for some ratio of axes, one of the modes will have a zero coefficient. This was tested out for a series of orientations near the *CT* and *DT* with the results shown in Fig. 1.16. Positive angle crystals had zero coefficients at ratios of axes varying from 1 to .855
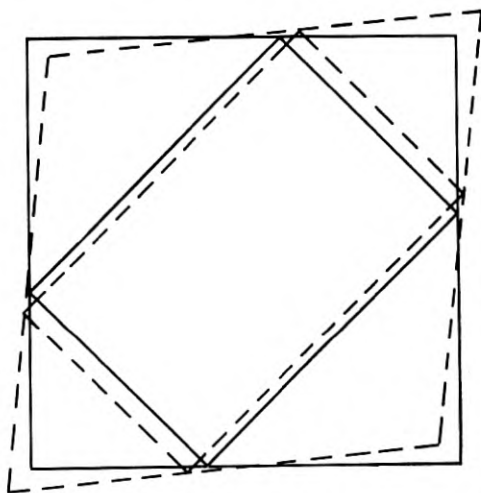
Fig. 1.15—Relation between a face shear mode and two coupled longitudinal modes

depending on the angle while negative angle crystals had zero coefficients at ratios from .64 to 1.0. For positive angle crystals it was the higher frequency mode that was the stronger and could be given the zero coefficient, while for the negative angle crystals it was the lower frequency mode that was the stronger and corresponded to the face shear mode.

Several of the positive angle crystals were measured over a temperature range with the results shown by Fig. 1.17. For angles above 51°-30′ the curvature was positive, while for angles below 51°-30′ the curvature was negative. Right at 51°-30′ the large square law curvature term disappeared and the frequency was constant to one part in a million over a 100-degree Centigrade range centered at 50°C. as shown by Fig. 1.18. Some further experiments showed that this flat range could be moved around a bit by

changing the angle of cut and the ratio of axes simultaneously. To go from −25°C. to +75°C. with a mid-range at 25°C., a crystal cut at 51°-7.5′
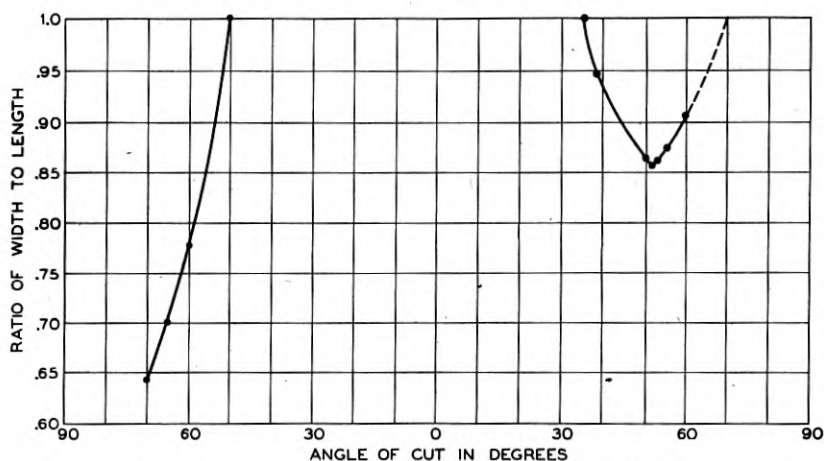
Fig. 1.16.—Relation between angle of cut and ratio of width to length for zero temperature coefficient for *G* type crystals
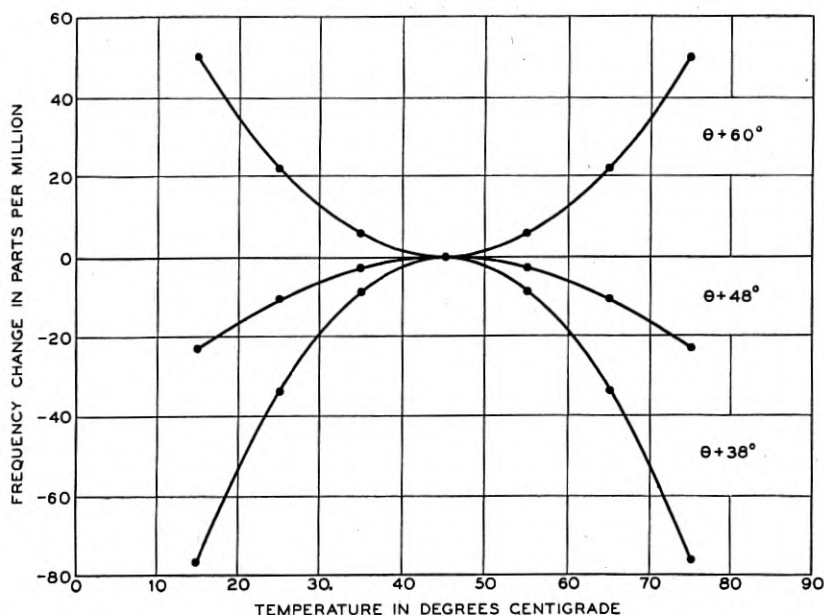
Fig. 1.17—Temperature frequency relations for various angles for *G* type crystals

with a ratio of axes of 0.859 is required. The *GT* crystal has been used quite extensively in frequency and time standards and in filters meeting rigid phase requirements.
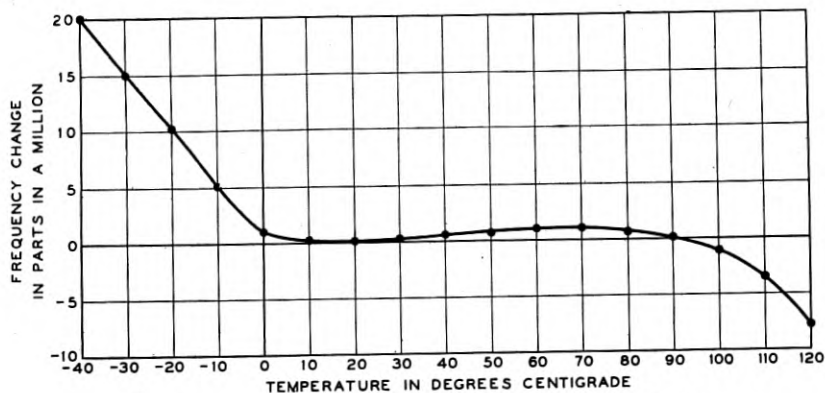
Fig. 1.18—Temperature frequency characteristic for *GT* crystal
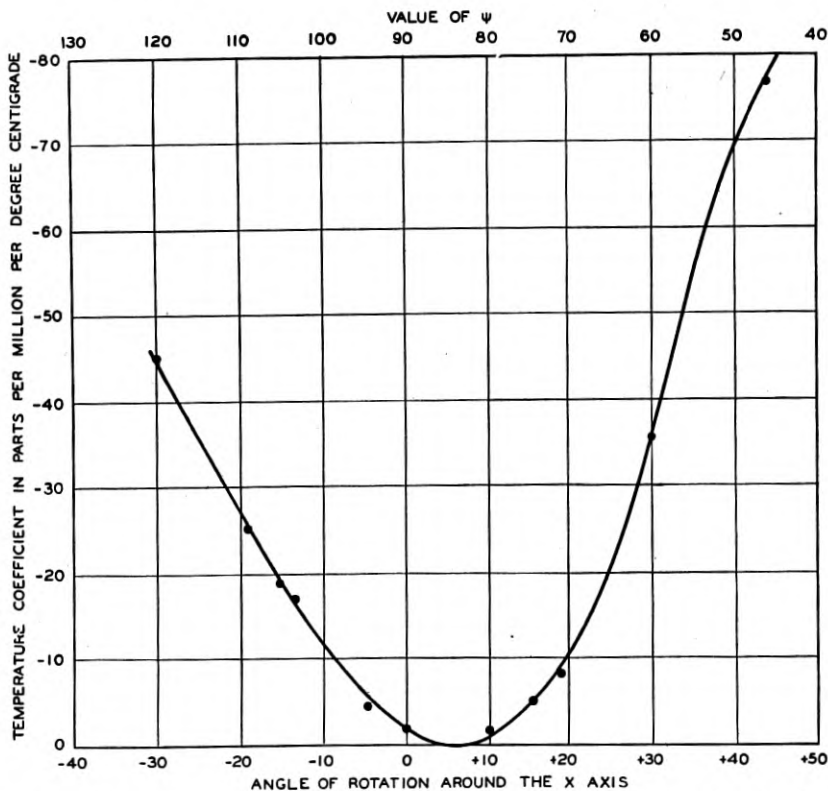
Fig. 1.19—Temperature coefficients of long thin rotated *X* cut crystals. Angle of rotation measured between length and *Y* axis. Dots are measured points. Solid line calculated from temperature coefficients evaluated in the appendix.

Two other cuts not previously described are shown also by Fig. 1.9. They are the *MT* low coefficient longitudinally vibrating crystal and the *NT* low coefficient flexurally vibrating crystal. Both of these are related to the $+5°$ *X* cut crystal of Fig. 1.9. As shown by Fig. 1.19 a long thin $5°$ *X* cut crystal is the best length direction for an *X* cut crystal to obtain a low-temperature coefficient. Figure 1.19 plots the temperature coefficients for long thin oriented *X* cut crystals, and this data is used in the appendix to derive the temperature coefficients of the six elastic constants. However, as the width of the crystal is increased the temperature coefficient becomes highly negative as shown by Fig. 1.20.
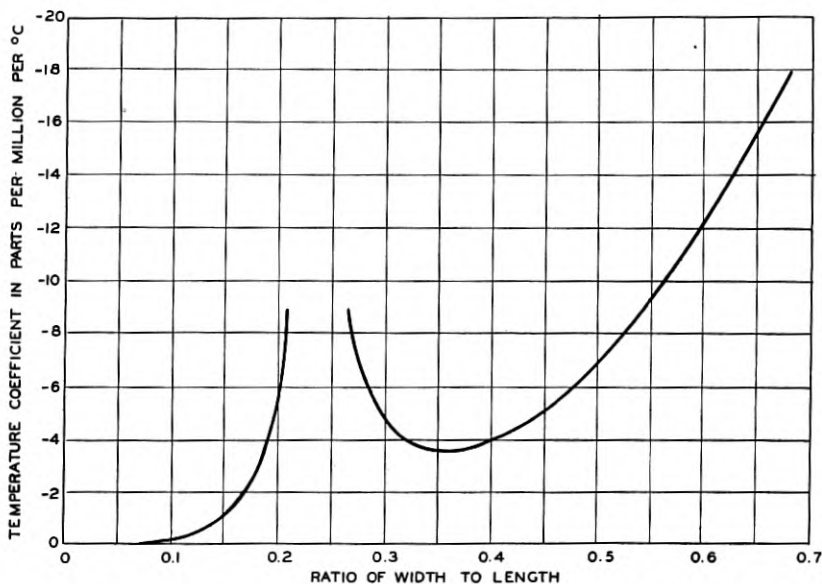
Fig. 1.20—Temperature coefficient of $a +5°$ *X* cut crystal ($\varphi = 0°$; $\theta = 90°$; $\psi = 85°$) as a function of the ratio of width to length. Ratio of thickness to length = 0.05.

This change of coefficient occurs due to the fact that as the crystal width is increased, the face shear mode of motion becomes more strongly excited and contributes to the elastic constant. Then since the temperature coefficient of the shear elastic constant is highly negative for this orientation the temperature coefficient of the $+5°$ *X* cut crystal becomes more highly negative as the width is increased.

The *MT* longitudinally vibrating crystal employs a rotation of the plane of the crystal cut about the *Y'* or length axis. The effect of this rotation is to change the temperature coefficient of the shear mode from highly negative to nearly zero. The result is that the temperature coefficient becomes

very low and nearly independent of the width length ratio. The NT low coefficient flexurally vibrating crystal is similar to the MT but requires a somewhat higher rotation about the Y' axis to produce a low coefficient.

The MT crystal has been used in narrow band filters such as pilot channel filters of the cable carrier system and in oscillators having frequencies between 50 kilocycles and 100 kilocycles. The NT flexure crystal is capable of producing frequencies as low as 4 kilocycles, and can be used to produce filters and control oscillators in the frequency range from 4 kilocycles to 50 kilocycles. Crystals of this type have been used with the Western Electric frequency modulation broadcast transmitter.[15] Operating in the region of 5 kilocycles, they maintain the frequency of the transmitter to $\pm.0025$ per cent without temperature regulation. These two crystals will be described in more detail in a subsequent paper.

APPENDIX A

VOIGT'S ELASTIC AND PIEZOELECTRIC RELATIONS AND THEIR APPLICATION TO THE DETERMINATION OF LOW TEMPERATURE COEFFICIENT CRYSTALS

A.1 MATHEMATICAL EXPRESSIONS FOR PIEZOELECTRIC RELATIONS

As mentioned in the historical introduction, Voigt formulated a mathematical relation between the stresses, strains, polarizations, and electric fields existing in a crystal. For a general crystal devoid of symmetry these relations take the form

$$-x_x = s_{11}^E X_x + s_{12}^E Y_y + s_{13}^E Z_z + s_{14}^E Y_z + s_{15}^E Z_x$$
$$+ s_{16}^E X_y - d_{11} E_x - d_{21} E_y - d_{31} E_z$$

$$-y_y = s_{21}^E X_x + s_{22}^E Y_y + s_{23}^E Z_z + s_{24}^E Y_z + s_{25}^E Z_x$$
$$+ s_{26}^E X_y - d_{12} E_x - d_{22} E_y - d_{32} E_z$$

$$-z_z = s_{31}^E X_x + s_{32}^E Y_y + s_{33}^E Z_z + s_{34}^E Y_z + s_{35}^E Z_x$$
$$+ s_{36}^E X_y - d_{13} E_x - d_{23} E_y - d_{33} E_z$$

$$-y_z = s_{41}^E X_x + s_{42}^E Y_y + s_{43}^E Z_z + s_{44}^E Y_z + s_{45}^E Z_x$$
$$+ s_{46}^E X_y - d_{14} E_x - d_{24} E_y - d_{34} E_z$$

$$(A.1)$$

[15] "A New Broadcast Transmitter Circuit Design for Frequency Modulation," J. F. Morrison, *Proc. I. R. E.*, Vol. 28, No. 10, Oct. 1940, pp. 444–449.

$$-z_x = s_{51}^E X_x + s_{52}^E Y_y + s_{53}^E Z_z + s_{54}^E Y_z + s_{55}^E Z_x$$
$$+ s_{56}^E X_y - d_{15} E_x - d_{25} E_y - d_{35} E_z$$
$$-x_y = s_{61}^E X_x + s_{62}^E Y_y + s_{63}^E Z_z + s_{64}^E Y_z + s_{65}^E Z_x$$
$$+ s_{66}^E X_y - d_{16} E_x - d_{26} E_y - d_{36} E_z$$
$$P_x = -d_{11} X_x - d_{12} Y_y - d_{13} Z_z - d_{14} Y_z - d_{15} Z_x - d_{16} X_y + \kappa_1^F E_x$$
$$P_y = -d_{21} X_x - d_{22} Y_y - d_{23} Z_z - d_{24} Y_z - d_{25} Z_x - d_{26} X_y + \kappa_2^F E_y$$
$$P_z = -d_{31} X_x - d_{32} Y_y - d_{33} Z_z - d_{34} Y_z - d_{35} Z_x - d_{36} X_y + \kappa_3^F E_z$$

where x_x, y_y, z_z are the three longitudinal strains, y_z, z_x, x_y the three shearing strains, X_x, Y_y, Z_z the three longitudinal stresses Y_z, Z_x, X_y the three shearing stresses; P_x, P_y, P_z the x, y and z components of the polarization, and E_x, E_y, E_z the x, y and z components of the electric field. s_{11}^E, $\cdots$, s_{66}^E are the 36 elastic compliances. The superscript E is added to show that they must be measured when the field E is zero or the crystal plated and short circuited. As shown from section C of this appendix they can be measured from the resonances of completely plated crystals. From the principle of conservation of energy it can be shown that there is the general relation between the elastic compliances

$$s_{ij}^E = s_{ji}^E \tag{A.2}$$

so that the greatest number of compliance moduli is 21. In equation (A.1) the d_{ij} are the piezoelectric constants measured by observing the proportionality between the strains and the applied fields in the absence of external stresses. κ_i^F are the "free" susceptibilities of the crystals in the three space directions measured in the absence of stress. The susceptibilities are related to the "free" dielectric constants K_i^F by the equation

$$K_i^F = 1 + 4\pi\kappa_i^F \tag{A.3}$$

In addition to these equations we have also that the charge per unit area Q on the surface is related to the field and polarization by

$$Q_x = \frac{E_x}{4\pi} + P_x$$
$$= \frac{E_x K_1^F}{4\pi} - d_{11} X_x - d_{12} Y_y - d_{13} Z_z - d_{14} Y_z - d_{15} Z_x - d_{16} X_y$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \tag{A.4}$$
$$Q_z = \frac{E_z}{4\pi} + P_z$$
$$= \frac{E_z K_3^F}{4\pi} - d_{31} X_x - d_{32} Y_y - d_{33} Z_z - d_{34} Y_z - d_{35} Z_x - d_{36} X_y$$

These equations hold for the most general type of crystal. In addition Voigt showed that if there was any symmetry existing in the crystal, a number of the constants were zero and certain relations existed between other constants. For example quartz has a trigonal symmetry about the Z or optic axis, and three digonal axes of symmetry (the three X or electrical axes) about which it is necessary to turn through an angle of $180°$ before the original pattern is restored. Voigt showed that by expressing the relations (A.1) in terms of rotated axes and imposing the symmetry conditions, the following relations existed between the elastic and piezoelectric coefficients

$$s_{15}^E = s_{16}^E = s_{25}^E = s_{26}^E = s_{34}^E = s_{35}^E = s_{36}^E = s_{45}^E = s_{46}^E = 0$$

$$s_{22}^E = s_{11}^E; \qquad s_{23}^E = s_{13}^E; \qquad s_{24}^E = -s_{14}^E; \qquad s_{44}^E = s_{55}^E;$$

$$s_{56}^E = 2s_{14}^E; \qquad s_{66}^E = 2(s_{11}^E - s_{12}^E)$$

$$d_{13} = d_{15} = d_{16} = d_{21} = d_{22} = d_{23} = d_{24} = d_{31} = d_{32} \qquad \text{(A.5)}$$

$$= d_{33} = d_{34} = d_{35} = d_{36} = 0$$

$$d_{12} = -d_{11}; \qquad d_{25} = -d_{14}; \qquad d_{26} = -2d_{11}$$

$$\kappa_1^F = \kappa_2^F$$

Hence the relations between the stresses, strains, polarizations and fields for quartz reduce to the simpler forms

$$-x_x = s_{11}^E X_x + s_{12}^E Y_y + s_{13} Z_z + s_{14}^E Y_z - d_{11} E_x$$

$$-y_y = s_{12}^E X_x + s_{11}^E Y_y + s_{13} Z_z - s_{14}^E Y_z + d_{11} E_x$$

$$-z_x = s_{13} X_x + s_{13} Y_y + s_{33} Z_z$$

$$-y_z = s_{14}^E X_x - s_{14}^E Y_y + s_{44}^E Y_z - d_{14} E_x$$

$$-z_x = s_{44}^E Z_x + 2s_{14}^E X_y + d_{14} E_y$$

$$-x_y = 2s_{14}^E Z_x + 2(s_{11}^E - s_{12}^E) X_y + 2d_{11} E_y \qquad \text{(A.6)}$$

$$Q_x = \frac{E_x K_1^F}{4\pi} - d_{11} X_x + d_{11} Y_y - d_{14} Y_z$$

$$Q_y = \frac{E_y K_1^F}{4\pi} + d_{14} Z_x + 2d_{11} X_y$$

$$Q_z = \frac{E_z K_3}{4\pi}$$

The superscripts have been left off the constants s_{13}, s_{33} and K_3 since it will be shown that their values are not affected by the way in which they are measured.

Equation (A.5) is not the only way of relating the elastic and electric quantities. For example if we substitute the values of the fields of the last three equations of (A.6) in the first six equations, we can write

$$-x_x = s_{11}^Q X_x + s_{12}^Q Y_y + s_{13} Z_z + s_{14}^Q Y_z - g_{11} Q_x$$

$$-y_y = s_{12}^Q X_x + s_{11}^Q Y_y + s_{13} Z_z - s_{14}^Q Y_z + g_{11} Q_x$$

$$-z_z = s_{13} X_x + s_{13} Y_y + s_{33} Z_z$$

$$-y_z = s_{14}^Q X_x - s_{14}^Q Y_y + s_{44}^Q Y_z - g_{14} Q_x$$

$$-z_x = s_{44}^Q Z_x + 2s_{14}^Q X_y + g_{14} Q_y$$

$$-x_y = 2s_{14}^Q Z_x + 2(s_{11}^Q - s_{12}^Q)X_y + 2g_{11} Q_y \qquad \text{(A.7)}$$

$$E_x = \frac{4\pi}{K_1^F} Q_x + g_{11} X_x - g_{11} Y_y + g_{14} Y_z$$

$$E_y = \frac{4\pi}{K_1^F} Q_y - g_{14} Z_x - 2g_{11} X_y$$

$$E_z = \frac{4\pi}{K_3} Q_z$$

where

$$s_{11}^Q = s_{11}^E - \frac{4\pi d_{11}^2}{K_1^F}; \qquad s_{12}^Q = s_{12}^E + \frac{4\pi d_{11}^2}{K_1^F}; \qquad s_{14}^Q = s_{14}^E - \frac{4\pi d_{11} \, d_{14}}{K_1^F}$$

$$s_{44}^Q = s_{44}^E - \frac{4\pi d_{14}^2}{K_1^F}; \qquad g_{11} = \frac{4\pi d_{11}}{K_1^F}; \qquad g_{14} = \frac{4\pi d_{14}}{K_1^F}$$

The superscript Q is added to show that these are the elastic compliances that will be measured when the free charge on the surface is zero. These elastic constants are the ones measured when an unplated crystal is put in an airgap holder with a large air-gap since then no charge can flow to the surface of the crystal. The difference between the zero field and zero charge elastic constants for quartz is less than 1 per cent. For rochelle salt, however, they may differ by a factor of 4. For rochelle salt the principal piezoelectric constant d_{14} and the "free" dielectric constant K_1^F vary widely in value and phase angle with variations in temperature and frequency, whereas the piezoelectric constant g_{14} which is proportional to the ratio of these two is nearly a constant for all frequencies and temperatures, so that the formulation of equation (A.7) is more advantageous than that of equation (A.6). For quartz, however, both forms are reasonably constant. Furthermore the elastic constants of equation (A.6) are those for a plated crystal which are usually the ones of interest for a crystal employed in an oscillator or filter. Hence this formulation has been used in this appendix.

Both the formulation of (A.6) and (A.7) can be expressed in terms of the strains rather than the stresses. Since these are useful forms and are used later in this appendix, they are given below. Equations (A.8) are obtained directly from equations (A.6) by solving them simultaneously to replace the strain by the stress, while equations (A.10) are obtained in the same way from equations (A.7).

$$-X_x = c_{11}^E x_x + c_{12}^E y_y + c_{13} z_z + c_{14}^E y_z - e_{11} E_x$$

$$-Y_y = c_{12}^E x_x + c_{11}^E y_y + c_{13} z_z - c_{14}^E y_z + e_{11} E_x$$

$$-Z_z = c_{13} x_x + c_{13} y_y + c_{33} z_z$$

$$-Y_z = c_{14}^E x_x - c_{14}^E y_y + c_{44}^E y_z - e_{14} E_x$$

$$-Z_x = c_{44}^E z_x + c_{14}^E x_y + e_{14} E_y$$

$$-X_y = c_{14}^E z_x + \left(\frac{c_{11}^E - c_{12}^E}{2}\right) x_y + e_{11} E_y \qquad \text{(A.8)}$$

$$Q_x = \frac{E_x}{4\pi} + P_x = \frac{E_x K_1^C}{4\pi} + e_{11} x_x - e_{11} y_y + e_{14} y_z$$

$$Q_y = \frac{E_y}{4\pi} + P_y = \frac{E_y K_1^C}{4\pi} - e_{14} z_x - e_{11} x_y$$

$$Q_z = \frac{E_z}{4\pi} + P_z = \frac{E_z K_3}{4\pi}$$

where the relations for the elastic constants are

$$2c_{11}^E = \frac{s_{33}^E}{\alpha} + \frac{s_{44}^E}{\beta}; \qquad 2c_{12}^E = \frac{s_{33}^E}{\alpha} - \frac{s_{44}^E}{\beta}; \qquad c_{13} = \frac{-s_{13}}{\alpha};$$

$$c_{14}^E = \frac{-s_{14}^E}{\beta}; \qquad c_{33} = \frac{s_{11}^E + s_{12}^E}{\alpha}; \qquad c_{44}^E = \frac{s_{11}^E - s_{12}^E}{\beta};$$

$$c_{66}^E = \frac{c_{11}^E - c_{12}^E}{2} = \frac{s_{44}^E}{2\beta}; \qquad \alpha = s_{33}(s_{11}^E + s_{12}^E) - 2s_{13}^2;$$

$$\beta = s_{44}^E(s_{11}^E - s_{12}^E) - 2s_{14}^{E\,2}.$$

Conversely we can also write the useful relation

$$2s_{11}^E = \frac{c_{33}^E}{\alpha'} + \frac{c_{44}^E}{\beta'}; \qquad 2s_{12}^E = \frac{c_{33}^E}{\alpha'} - \frac{c_{44}^E}{\beta'};$$

$$s_{13} = \frac{-c_{13}}{\alpha'}; \qquad s_{14}^E = \frac{-c_{14}^E}{\beta'}; \qquad s_{33} = \frac{c_{11}^E + c_{12}^E}{\alpha'};$$

$$s_{44}^E = \frac{c_{11}^E - c_{12}^E}{\beta'}; \qquad s_{66}^E = 2(s_{11}^E - s_{12}^E) = \frac{2c_{44}^E}{\beta'};$$

$$\alpha' = c_{33}(c_{11}^E + c_{12}^E) - 2c_{13}^2; \qquad \beta' = c_{44}^E(c_{11}^E - c_{12}^E) - 2c_{14}^{E\,2}.$$

For the piezoelectric constants

$$e_{11} = d_{11}(c^E_{11} - c^E_{12}) + d_{14}c^E_{14} ; \qquad e_{14} = 2d_{11}c^E_{14} + d_{14}c^E_{44} ;$$

and conversely

$$-d_{11} = e_{11}(s^E_{11} - s^E_{12}) + e_{14}s^E_{14} ; \qquad -d_{14} = 2e_{11}s^E_{14} + e_{14}s^E_{44} .$$

The dielectric constant K^C_1 denotes the clamped dielectric constant, i.e., the constant measured when the crystal is free from strain. This is related to the free dielectric constant K^F_1 by the equation

$$K^C_1 = K^F_1 - 4\pi[d_{14}e_{14} + 2d_{11}e_{11}]. \tag{A.9}$$

In a similar way if we solve equations (A.7) simultaneously, for the stresses in terms of the strains, we have

$$-X_x = c^Q_{11}x_x + c^Q_{12}y_y + c_{13}z_z + c^Q_{14}y_z - f_{11}Q_x ;$$

$$-Y_y = c^Q_{12}x_x + c^Q_{11}y_y + c_{13}z_z - c^Q_{14}y_z + f_{11}Q_x ;$$

$$-Z_z = c_{13}x_x + c_{13}y_y + c_{33}z_z ;$$

$$-Y_z = c^Q_{14}x_x - c^Q_{14}y_y + c^Q_{44}y_z - f_{14}Q_x ;$$

$$-Z_x = c^Q_{44}z_x + c^Q_{14}x_y + f_{14}Q_y ;$$

$$-X_y = c^Q_{14}z_x + \left(\frac{c^Q_{11} - c^Q_{12}}{2}\right)x_y + f_{11}Q_y ; \tag{A.10}$$

$$E_x = \frac{4\pi}{K^C_1}Q_x - f_{11}x_x + f_{11}y_y - f_{14}y_z ;$$

$$E_y = \frac{4\pi}{K^C_1}Q_y + f_{14}z_z + f_{11}x_y ;$$

$$E_z = \frac{4\pi}{K_3}Q_z .$$

where the c^Q constants are related to the s^Q constants as in equation (A.8). The piezoelectric relations are

$$f_{11} = g_{11}(c^Q_{11} - c^Q_{12}) + g_{14}c^Q_{14} ; \qquad f_{14} = 2g_{11}c^Q_{14} + g_{14}c^Q_{44} ;$$

or conversely

$$-g_{11} = f_{11}(s^Q_{11} - s^Q_{12}) + e_{14}s^Q_{14} ; \qquad -g_{14} = 2f_{11}s^Q_{14} + f_{14}s^Q_{44} ;$$

while the dielectric relation between the free and clamped crystal

$$\frac{4\pi}{K^F_1} = \frac{4\pi}{K^C_1} - (g_{14}f_{14} + 2g_{11}f_{11}). \tag{A.11}$$

Equations (A.10) might also have been obtained directly from equations (A.8) by substituting the charges from the last three equations in terms of the fields. This substitution yields the additional relations

$$c_{11}^Q = c_{11}^E + \frac{4\pi e_{11}^2}{K_1^C}; \quad c_{12}^Q = c_{12}^E - \frac{4\pi e_{11}^2}{K_1^C}; \quad c_{13}^Q = c_{13}^E;$$

$$c_{14}^Q = c_{14}^E + \frac{4\pi e_{11} e_{14}}{K_1^C}; \quad c_{33}^Q = c_{33}^E;$$

$$c_{44}^Q = c_{44}^E + \frac{4\pi e_{14}^2}{K_1^C}; \quad c_{66}^Q = \frac{c_{11}^Q - c_{12}^Q}{2} = \frac{c_{11}^E - c_{12}^E + \frac{8\pi e_{11}^2}{K_1^C}}{2}; \quad (A.12)$$

$$f_{11} = \frac{4\pi}{K_1^C} e_{11}; \quad f_{14} = \frac{4\pi}{K_1^C} e_{14}.$$

(A.2). VALUES OF THE ELASTIC AND PIEZOELECTRIC CONSTANTS

The first and one of the best determinations of the elastic constants of quartz was made by Voigt. Using static deformations of unplated crystals he determined the elastic constants to be

$$c_{11} = 85.1 \times 10^{10} \text{ dynes/cm}^2; \quad c_{12} = 6.95 \times 10^{10};$$

$$c_{13} = 14.1 \times 10^{10}; \quad c_{14} = 16.8 \times 10^{10}$$

$$c_{33} = 105.3 \times 10^{10}; \quad c_{44} = 57.1 \times 10^{10}; \quad (A.13)$$

$$c_{66} = \left(\frac{c_{11} - c_{12}}{2}\right) = 39.1 \times 10^{10}$$

From these the moduli of compliance can be calculated and are

$$s_{11} = 129.8 \times 10^{-14} \text{ cm}^2/\text{dyne}; \quad s_{12} = -16.6 \times 10^{-14};$$

$$s_{13} = -15.2 \times 10^{-14}; \quad s_{14} = -43.1 \times 10^{-14};$$

$$s_{33} = 99.0 \times 10^{-14}; \quad s_{44} = 200.5 \times 10^{-14}; \quad (A.14)$$

$$s_{66} = 2(s_{11} - s_{12}) = 292.8 \times 10^{-14}.$$

Whether these are zero field or zero charge constants is not known. If they were measured in a room with high humidity, the polarization produced by strain would soon be annulled by a current flow through the leakage resistance of the adsorbed moisture, and the constants would be c_{ij}^E or s_{ij}^E. On the other hand if the displacements were measured in a very dry room, the leakage resistance is very small and it may take hours to annul the polarization through a leakage current flow. In that case the constants measured

would be c_{ij}^Q or s_{ij}^Q . In any case the difference was probably less than the accuracy of measurement.

Later measurements by Perrier and Mandrot for two of the constants s_{11} and s_{33} give the values

$$s_{11} = 127.3 \times 10^{-14}; \qquad s_{33} = 97 \times 10^{-14} \qquad (A.15)$$

By using the measured resonance frequencies of known modes of motion, the uncertainty of the type of elastic constant can be removed, for the alternations occur so fast that the leakage resistance has little effect. If a crystal is lightly plated, it is shown in the next section that the resonant frequency of a length vibrating bar will be determined by the zero field elastic constants s_{ij}^E . On the other hand if an unplated crystal is measured in an air gap holder with a large air gap it has been shown that[1] the frequency measured will be determined by the zero charge elastic constants s_{ij}^Q or c_{ij}^Q . A careful measurement of the elastic constants of quartz has recently been made by Atanasoff and Hart[2]. Using thickness modes for

[1] The resonances of length vibrating crystals have been discussed by Cady, "The Piezo-electric Resonator and The Effect of Electrode Spacing on Frequency," *Physics*, Vol. 7, No. 7, July 1936, pages 237–259; and by the writer, "Dynamic Measurement of The Constants of Rochelle Salt," *Phys. Rev.*, Vol. 55, pages 775–789, April 15, 1939; while the resonances of thickness vibrating crystals have been discussed by Cady (above paper) and Lawson "The Vibration of Piezoelectric Plates," *Phys. Rev.*, Vol. 62, July 1, 1942, pp. 71–76. For a length vibrating crystal Cady shows that the resonant frequency for no air gap (plated crystal) is controlled by the constant $1/s_{11}^E$. For a crystal with a large air gap, the frequency is controlled by the constant.

$$1/s_{11}^E + 4\pi\, d_{11}^2/K_1^F\, s_{11}^{E\,2} = 1/s_{11}^Q .$$

Starting with equations of the form (A.10), the writer showed that the frequency of a bar in an air gap holder would be controlled by the constant $1/s_{11}^Q$, while the frequency of a plated crystal is determined by

$$s_{11}^Q \left/ \left(1 - \frac{d_{11}^2\, 4\pi}{K_1^C\, s_{11}^Q}\right) \right. = s_{11}^E .$$

For a thickness vibrating crystal for which the field is applied in the direction of wave propagation, Cady and Lawson find that the resonant frequency is controlled by the elastic constant

$$c^* = c_{11}^E + \frac{4\pi e_{11}^2}{K_1^C}\left[1 - \frac{8}{\pi^2\left[1 + K_1^C\left(\dfrac{D}{t} - 1\right)\right]}\right]$$

where D is the total separation between electrodes and t the thickness of the crystal. When the separation is infinite, the controlling elastic constant is $c_{11}^E + 4\pi e_{11}^2/K_1^C$ which, from equation (A.12) is c_{11}^Q. When the air gap is zero or $D = t$, the controlling constant is

$$c_{11}^E + \frac{4\pi e_{11}^2}{K_1^C}\left(1 - \frac{8}{\pi^2}\right)$$

which, for all practical purposes, can be taken as c_{11}^E for quartz.

[2] "Dynamical Determination of the Elastic Constants and their Temperature Coefficients for Quartz," *Phys. Rev.*, Vol. 59, No. 1 (85–96), Jan. 1, 1941.

relatively thick pieces of quartz, and determining the asymptotic value for high order harmonics, they obtained the elastic constants

$$c_{11} = 87.55 \times 10^{10} \text{ dynes/cm}^2; \quad c_{12} = 6.07 \times 10^{10}; \quad c_{13} = 13.3 \times 10^{10};$$

$$c_{14} = -c_{24} = 17.25 \times 10^{10}; \quad c_{33} = 106.8 \times 10^{10}; \quad c_{44} = 57.19 \times 10^{10}.$$

$$(A.16)$$

In addition they came to the conclusion that c_{56} had a value of 18.4×10^{10}, which was different from the value of c_{14} as required by theory. Their measurements were made with high harmonics in air gap holders so that the values measured should determine the c_{ij}^Q constant. To explain the discrepancy found, Lawson[3] has suggested that the c_{ij}^Q constants

$$c_{ij}^Q = c_{ij}^E + 4\pi e_{1i} e_{1j} / K_1^C \tag{A.17}$$

do not obey the same symmetry relations as the c_{ij}^E constants. This suggestion does not seem to be borne out by equations (A.10), from which the symmetry relations of the c_{ij}^Q constants can be determined. If we start with a generalized form of these equations applicable to any crystal and apply the symmetry relations for quartz, we find that it is still necessary to satisfy the symmetry relations between the constants found previously and in particular

$$c_{56}^Q = c_{14}^Q \tag{A.18}$$

In order to investigate this matter further, and to obtain more reliable values of the elastic constants, an analysis has been made of a number of measurements previously obtained for oriented crystals. In particular two families of oriented crystals were investigated. One family was a set of oriented X cut crystals vibrating longitudinally. They were cut with their major faces normal to the X axis and with their lengths at angles A_2 of from $+43°$ to $-79°$ with respect to the Y or mechanical axis. They were oriented similarly to the $+5°$ and $-18.5°$ filter crystals shown by Fig. 1.9. When these crystals are 7 to 10 times as long as they are wide or thick it has been shown previously[4] that their length resonances are determined very accurately by the equation

$$f_R = \frac{1}{2\ell_y} \sqrt{\frac{1}{\rho s_{22'}^E}} \tag{A.19}$$

where ℓ_y is the length of the crystal, ρ the density and $s_{22'}^E$ the inverse of Young's Modulus along the length for a plated crystal. This is related to the angle of cut A_2 by the equation

$$s_{22'}^E = s_{11}^E \cos^4 A_2 + s_{33}^E \sin^4 A_2 + 2s_{14}^E \cos^3 A_2 \sin A_2$$
$$+ (2s_{13} + s_{44}^E) \sin^2 A_2 \cos^2 A_2 \tag{A.20}$$

[3] A. W. Lawson, *Phys. Rev.*, 59, 838 (1941).
[4] "Electrical Wave Filters Employing Quartz Crystals as Elements," W. P. Mason, *B. S. T. J.*, Vol. XIII, pp. 405–452, July 1934. See Figs. 25, 31 and 32.

Since the resonant frequency of the plated crystal was measured, it was the zero potential elastic constant that was determined. These crystals were lightly plated with aluminum and it had been previously shown that the added plating would affect the frequency by considerably less than 0.1 per cent. The crystal orientations, their dimensions, the frequency constants and the values of $s_{22'}^E$ are shown by Table I.

These measured variations satisfy equation (A.20) for the variation of $s_{22'}^E$ with angle very well if we take

$$s_{11}^E = 127.9 \times 10^{-14} \text{ cm}^2/\text{dyne}; \quad s_{33} = 95.6 \times 10^{-14};$$

$$s_{14}^E = -44.6 \times 10^{-14}; \tag{A.21}$$

$$s_{44}^E + 2s_{13} = 175.8 \times 10^{-14}.$$

TABLE I

Angle of Cut, A_2	Dimension, mm			Resonant Frequency 25°C	Frequency Constant KC cms	Value of $s_{22'}^E$
	Length	Width	Thickness			
−79.5°	24.03	2.50	.502	130,700	314.1	$95.6 \times 10^{-14} \text{ cm}^2/\text{dyne}$
−18.5°	20.00	2.50	.502	127,710	255.4	144.5
−13.14°	19.99	2.97	.505	128,390	256.8	143.0
−12.33	19.98	2.95	.500	128,590	257.0	142.7
− 5.6	20.02	2.92	.500	132,130	264.5	134.7
− 1.4	20.03	3.02	.502	134,050	269.2	130.2
− .9°	19.97	2.99	.502	135,240	270.5	128.9
+ .36°	20.03	3.03	.508	135,890	272.0	127.5
+ .54	19.96	2.98	.506	135,920	272.1	127.3
+ 1.44	20.02	2.98	.505	136,890	274.0	125.7
+ 2.61	19.97	3.00	.505	138,400	276.5	123.3
+ 4.05	19.95	2.97	.510	139,900	279.0	121.2
+11.8	19.11	3.01	.500	154,600	295.4	108.1
+18.0	20.02	2.95	.500	155,380	311.1	97.5
+42.6	20.00	2.95	.500	174,750	349.5	77.25

This gives three of the constants directly, and a relation between two more. To obtain the remaining constants and to test out the hypothesis that there are seven elastic constants rather than six, use has been made of measurements made for thickness vibrating shear crystals obtained by rotating one edge about the X axis. These are the AT and BT series shown by Fig. 1.9. As shown by a former paper[5], the frequency of such crystals depends on the edge dimensions as well as the thickness dimensions. However, as the edge dimensions become large compared to the thickness dimension the principal frequency approaches an asymptotic value which is taken as that for the infinite plate. For the AT, BT and Y cut crystals these asymptotic values have been determined to have the values shown by Table II.

[5] "Low Temperature Coefficient Quartz Crystals," *B. S. T. J.*, Vol. XIX, pp. 74–93, Jan. 1940. See Fig. 5.

If we make the assumption that there are seven elastic constants and c_{56}^E differs from c_{14}^E, the frequency of this series of crystals will be[5]

$$f = \frac{1}{2\ell} \sqrt{\frac{c_{66'}^E}{\rho}} \quad \text{where} \quad c_{66'}^E = c_{66}^E \cos^2 A_2 + c_{44}^E \sin^2 A_2 - c_{56}^E \sin 2A_2 \quad \text{(A.22)}$$

The determination for the Y cut gives directly

$$c_{66}^E = 40.5 \times 10^{10} \text{ dynes per square cm.} \quad \text{(A.23)}$$

The other two cuts give the values

$$c_{56}^E = 18.2 \times 10^{10}; \qquad c_{44}^E = 58.65 \times 10^{10} \quad \text{(A.24)}$$

To test out the hypothesis that c_{56}^E differs from c_{14}^E or s_{56}^E from $2s_{14}^E$ we can make use of equation (A.8) writing c_{56}^E in place of c_{14}^E. Then solving these equations simultaneously we find

$$s_{44}^E = \frac{c_{66}^E}{(c_{44}^E c_{66}^E - c_{56}^{E2})}; \quad s_{56}^E = \frac{-c_{56}^E}{(c_{44}^E c_{66}^E - c_{56}^{E2})}; \quad s_{66}^E = \frac{c_{44}^E}{(c_{44}^E c_{66}^E - c_{56}^{E2})} \quad \text{(A.25)}$$

TABLE II

Crystal	Angle of Cut A_2	Asymytotic Frequency constant KC mms	Value of $c_{66'}^E$
AT	$+35°\ 15'$	1661.5	29.39×10^{10} dynes/cm²
Y Cut	0	1954	40.50
BT	$-49°$	2549	68.86

Substituting in the values from (A.23) and (A.24) we find

$$s_{44}^E = 197.8 \times 10^{-14} \text{ cm}^2/\text{dyne}; \qquad s_{56}^E = -89.0 \times 10^{-14};$$
$$s_{66}^E = 2(s_{11}^E - s_{12}^E) = 286.5 \times 10^{-14}. \quad \text{(A.26)}$$

Comparing the value of s_{56}^E with $2s_{14}^E$ given in equation (A.21) we see that they are equal within the experimental error, so that these measurements do not indicate that there are seven elastic constants but only the customary six. Using these values all the elastic constants can be evaluated as shown by Table III.

Measurements have also been made to determine accurately the piezoelectric constants. This was done by using the ratios of capacities of two standard rotated X cut crystals for which these ratios have been accurately determined. As shown by section C of this appendix, the ratio of capacities r of a crystal is related to the piezoelectric constant d_{12}', the elastic constant $s_{22'}^E$ and the free dielectric constant K_1^F by the equation

$$r = \text{ratio of capacities} = \frac{\pi^2}{8}\left(\frac{1 - k^2}{k^2}\right) \quad \text{(A.27)}$$

where k the electromechanical coupling is given by

$$k = d'_{12} \sqrt{\frac{4\pi}{K_1^F s_{22}^{E'}}}. \tag{A.28}$$

The two crystal cuts and their constants are given in Table IV.
Only the numerical value and not the sign are determined for d'_{12}.

<div style="text-align:center">TABLE III</div>

Elastic Compliance Moduli	Elastic Stiffness Moduli
$s_{11}^E = 127.9 \times 10^{-14}$ cm²/dyne	$c_{11}^E = 86.05 \times 10^{10}$ dynes/cm²
$s_{12}^E = -15.35$	$c_{12}^E = 5.05$
$s_{13} = -11.0$	$c_{13} = 10.45$
$s_{14}^E = -44.6$	$c_{14} = 18.25$
$s_{33} = 95.6$	$c_{33} = 107.1$
$s_{44}^E = 197.8$	$c_{44} = 58.65$
$s_{66}^E = 2(s_{11}^E - s_{12}^E) = 286.5$	$c_{66}^E = \dfrac{c_{11}^E - c_{12}^E}{2} = 40.5$

<div style="text-align:center">TABLE IV</div>

Angle of Cut, A_2	Ratio of Capacities	Value of $s_{22}^{E'}$	Value of K_1^F	Value of d'_{12}
$-18.5°$ X cut	137	144.5	4.58	6.85×10^{-8}
$0°$ X cut	125	127.9	4.58	6.76×10^{-8}

<div style="text-align:center">TABLE V</div>

Piezoelectric constant	Value in cgs electrostatic units	Piezoelectric constant	Value in cgs electrostatic units
d_{11}	-6.76×10^{-8}	g_{11}	-18.55×10^{-8}
d_{14}	2.56×10^{-8}	g_{14}	7.02×10^{-8}
e_{11}	-5.01×10^{4}	f_{11}	-13.85×10^{4}
e_{14}	$-.97 \times 10^{4}$	f_{14}	-2.68×10^{4}

The variation of d'_{12} as a position of angle has been shown to be[6]

$$d'_{12} = -\tfrac{1}{2}[d_{11}(1 + \cos 2A_2) + d_{14} \sin 2A_2] \tag{A.29}$$

The two values of d'_{12} of table IV are satisfied by

$$d_{11} = -6.76 \times 10^{-8}; \qquad d_{14} = +2.56 \times 10^{-8} \tag{A.30}$$

[6] See "Electrical Wave Filters Employing Quartz Crystals as Elements," W. P. Mason, B. S. T. J., Vol. XIII, 405 (July 1934).

From these values and the elastic constants of Table III we can calculate all the different forms of the piezoelectric constants. These are given in Table V.

(A.3). DERIVATION OF EQUIVALENT CIRCUIT OF CRYSTAL

The electrical impedance and electrical equivalent circuit for a fully plated crystal can be derived from the piezoelectric relations of equation (A.6) taken together with Newton's law of motion

$$F_y = ma = (\rho \, dx \, dy \, dz) \frac{d^2\xi}{dt^2} \tag{A.31}$$

where m is mass of an elementary volume $dx \, dy \, dz$, a the acceleration, and ξ is the displacement of the element in the y direction. If we consider a long thin X cut crystal with its length in the y direction, the piezoelectric relations of interest are

$$-y_y = s_{12}^E X_x + s_{11}^E Y_y + s_{13} Z_z - s_{14}^E Y_z + d_{11} E_x;$$
$$Q_x = \frac{E_x K_1^F}{4\pi} - d_{11} X_x + d_{11} Y_y - d_{14} Y_z. \tag{A.32}$$

For a long thin crystal with its long dimension in the Y direction we can set

$$X_x = Z_z = Y_z = 0 \tag{A.33}$$

This follows since the crystal is free from external forces, and hence these stresses on the edges of the crystal must be zero. On account of the small x and z dimensions, the rate of change of these stresses with x or z will have to be high in order that the stresses shall differ appreciably from zero, and there are no mechanical strains causing a high stress gradient. Then for a long thin bar the piezoelectric equations can be written

$$-y_y = s_{11}^E Y_y + d_{11} E_x;$$
$$Q_x = \frac{E_x K_1^F}{4\pi} + d_{11} Y_y. \tag{A.34}$$

Let us next consider a small cross section of the crystal with a dimension dy along the crystal length. The total force on the section is a resultant of the difference in stresses on the two faces or equal to

$$\ell_w \ell_t [Y_{y_1} - Y_{y_2}] = -\ell_w \ell_t \frac{\partial Y_y}{\partial y} \, dy = F_y \tag{A.35}$$

where Y_y the stress is considered as a compressional force acting on the faces of the element. By Newton's law of motion (A.31) we have

$$-\ell_w \ell_t \, dy \frac{\partial Y_y}{\partial y} = \ell_w \ell_t \, dy \, \rho \frac{d^2\xi}{dt^2} \quad \text{or} \quad \frac{\partial Y_y}{\partial y} = -\rho \frac{d^2\xi}{dt^2} \tag{A.36}$$

For a completely plated crystal such as we are considering, the potential gradient E_x will be independent of the y direction, since any charge distribution will be equalized with the speed of light which is much higher than the speed of sound in the crystal. Then equation (A.34) when differentiated by y becomes

$$-\frac{\partial Y_y}{\partial y} = -\frac{\partial^2 \xi}{\partial y^2} = s_{11}^E \frac{\partial y_y}{\partial y}. \tag{A.37}$$

Introducing equation (A.36), the equation of motion for a plated crystal becomes

$$\frac{\partial^2 \xi}{\partial y^2} = s_{11}^E \rho \frac{d^2 \xi}{dt^2}. \tag{A.38}$$

For simple harmonic motion the variation of ξ with time can be written in the usual form

$$\xi = \xi e^{j\omega t}, \tag{A.39}$$

so that for simple harmonic motion equation (A.38) becomes

$$\frac{d^2 \xi}{dy^2} - \omega^2 s_{11}^E \rho \xi = \frac{d^2 \xi}{dy^2} - \frac{\omega^2}{v^2} \xi = 0 \tag{A.40}$$

where v the velocity of sound in the plated crystal is given by the formula

$$v^2 = \frac{1}{\rho s_{11}^E}. \tag{A.41}$$

A solution of equation (A.40) with two arbitrary boundary conditions is

$$\xi = A \cos \frac{\omega}{v} y + B \sin \frac{\omega y}{v}. \tag{A.42}$$

To determine the constants A and B, use is made of equation (A.34). Differentiating (A.42)

$$-\frac{\partial \xi}{\partial y} = -y_y = \frac{\omega}{v} \left[A \sin \frac{\omega}{v} y - B \cos \frac{\omega y}{v} \right] = s_{11}^E Y_y + d_{11} E_x. \tag{A.43}$$

When $y = 0$ and $y = \ell$ the bar length

$$Y_y = Y_{y_1} \quad \text{and} \quad Y_y = Y_{y_2} \tag{A.44}$$

provided the crystal is driving a load. For most electrical cases the only load driven is an air load and this is usually very small so that it is customary to set $Y_{y_1} = Y_{y_2} = 0$. Under these conditions

$$-\frac{\omega}{v} B = d_{11} E_x \quad \text{and} \quad \frac{\omega}{v} \left[A \sin \frac{\omega \ell}{v} - B \cos \frac{\omega \ell}{v} \right] = d_{11} E_x. \tag{A.45}$$

Solving these equations for A and B and substituting in (A.43) we have

$$-y_y = d_{11} E_x \left[\tan \frac{\omega \ell}{2v} \sin \frac{\omega y}{v} + \cos \frac{\omega y}{v} \right] = s_{11}^E Y_y + d_{11} E_x$$

or $\qquad Y_y = -\frac{d_{11} E_x}{s_{11}^E} \left[1 - \dfrac{\cos \dfrac{\omega(y - \ell/2)}{v}}{\cos \dfrac{\omega \ell}{2v}} \right].$ $\qquad$ (A.46)

The electrical impedance measured at the terminals of a plated crystal is then determined by substituting the value of Y_y in the last of equations (A.34) and integrating the charge Q over the whole surface. The current into the crystal is then

$$i = j\omega Q = j\omega \ell_w \int_0^\ell E_x \left[\frac{K_1^F}{4\pi} - \frac{d_{11}^2}{s_{11}^E} \left(1 - \frac{\cos \omega(y - \ell/2)}{\cos \dfrac{\omega \ell}{2v}} \right) \right] dy$$

$$= j\omega E_x \ell_w \ell \left[\frac{K_1^F}{4\pi} - \frac{d_{11}^2}{s_{11}^E} \left(1 - \frac{\tan \dfrac{\omega \ell}{2v}}{\dfrac{\omega \ell}{2v}} \right) \right] \qquad (A.47)$$

$$= j\omega E_x \ell \ell \left[\frac{K_1^{LC}}{4\pi} + \frac{d_{11}^2}{s_{11}^E} \frac{\tan \dfrac{\omega \ell}{2v}}{\dfrac{\omega \ell}{2v}} \right]$$

where $K_1^{LC} = K_1^F - \dfrac{4\pi d_{11}^2}{s_{11}^E}$ is called the longitudinally clamped dielectric constant, i.e. the dielectric constant that would be measured if we suppress the longitudinal strain along the y axis but not the other strains. The admittance of the crystal then is

$$\frac{i}{E} = \frac{i}{E_x \ell_t} = \frac{j\omega \ell_w \ell}{\ell_t} \left[\frac{K_1^{LC}}{4\pi} + \frac{d_{11}^2}{s_{11}^E} \frac{\tan \dfrac{\omega \ell}{2v}}{\dfrac{\omega \ell}{2v}} \right]. \qquad (A.48)$$

This consists of two terms which represent parallel branches in the equivalent circuit. One of these is the capacitance

$$C_0 = \frac{\ell_w \ell K_1^{LC}}{4\pi \ell_t} \text{ cgs units} = \frac{\ell_w \ell K_1^{LC}}{4\pi \ell_t 9 \times 10^{11}} \text{ farads} \qquad (A.49)$$

The other branch contains the impedance

$$\frac{-jl_t}{\omega l_w \ell}\left[s_{11}^E \frac{\frac{\omega \ell}{2v}}{d_{11}^{2}\tan\frac{\omega\ell}{2v}}\right]\text{cgs units} = \frac{-jl_t s_{11}^E \frac{\omega\ell}{2v}\times 9\times 10^{11}}{\omega l_w \ell\, d_{11}^{2}\tan\frac{\omega\ell}{2v}}\text{ohms}\quad\text{(A.50)}$$

This branch will have a zero impedance or will resonate when the tangent is infinite or when

$$\frac{2\pi f_R \ell}{2v} = \frac{\pi}{2}\text{ or } f_R = \frac{v}{2\ell} = \frac{1}{2\ell\sqrt{\rho s_{11}^E}}\quad\text{(A.51)}$$

Hence for a fully plated crystal it is the zero field elastic constant that determines the resonant frequency.

Near this resonant frequency, the impedance of equation (A.50) can be represented by a series capacitance and inductance having the values

$$C_1 = \frac{l_w \ell}{l_t}\frac{8}{\pi^2}\frac{d_{11}^{2}}{s_{11}^E\times 9\times 10^{11}},\qquad L_1 = \frac{\rho s_{11}^{E^2}\ell l_t\times 9\times 10^{11}}{8 l_w\, d_{11}^{2}}\quad\text{(A.52)}$$

Taking the ratio between C_0 and C_1 we have

$$\frac{C_0}{C_1} = r = \frac{\pi^2}{8}\left(\frac{K_1^{LC} s_{11}^E}{4\pi\, d_{11}^{2}}\right) = \frac{\pi^2}{8}\left(\frac{1 - \frac{4\pi\, d_{11}^{2}}{s_{11}^E K_1^F}}{\frac{4\pi\, d_{11}^{2}}{K_1^F s_{11}^E}}\right) = \frac{\pi^2}{8}\left(\frac{1 - k^2}{k^2}\right)\quad\text{(A.53)}$$

where k the coefficient of electromechanical coupling is equal to

$$k = d_{11}\sqrt{\frac{4\pi}{K_1^F s_{11}^E}}.\quad\text{(A.54)}$$

These values are used in equations (A.27) and (A.28) to evaluate the piezo-electric constants of quartz.

A.4. Use of Voigt's Relations in Locating Regions of Low Temperature Coefficient Crystals for Simple Modes of Motion

In Section 1.5 of the text, the statement is made that all longitudinally vibrating crystals of quartz have a zero or negative temperature coefficients. This can be proved from Voigt's relations for quartz and a knowledge of the temperature coefficients of the six elastic constants of quartz. Since the same method can be used to locate the regions of low temperature coefficient for other simple modes of motion a short discussion of the method is given here.

The Voigt relations given in equation (A.6) give the values of the piezo-

electric and elastic constants for crystals with their three edge dimensions along the three crystallographic axes. Most low-coefficient crystals, however, are oriented crystals with one or more of their edges lying along directions not parallel to the crystallographic axes. The theory of elasticity, however, provides methods for calculating the values of the constants for rotated axes. If the rotated axes X', Y', Z' are related to the crystallographic axes X, Y, and Z by the relation

$$
\begin{array}{c|ccc}
 & X & Y & Z \\
\hline
X' & \ell_1 & m_1 & n_1 \\
Y' & \ell_2 & m_2 & n_2 \\
Z' & \ell_3 & m_3 & n_3
\end{array}
\tag{A.55}
$$

where ℓ_1, $\cdots$, n_3 are the direction cosines between the axes indicated, the theory of elasticity provided relations between the stresses of the rotated axes and the stresses of the crystallographic axes, between the strains of the rotated axes and the strains of the crystallographic axes, and between the field, polarizations, or charges of the rotated axes and the same quantities for the crystallographic axes. Then if we express[7] the relation between the stress, strain and fields for the rotated axes, the elastic and piezoelectric constants are determined.

Two shorthand methods are also available for calculating the constants of rotated crystals. One method[8] is the matrix method which is based upon the fact that relations in (A.6) can be expressed in a matrix equation

$$
-\epsilon = s^E X + dE
\tag{A.56}
$$

where ϵ are the strain components, X the stress components, s^E the elastic compliance matrix, d the piezoelectric matrix and E the field components. By applying the rules of matrix multiplication the s and d matrices can be transformed to rotated axes having the direction cosines of equation (A.57) with respect to the crystallographic axes. The other method is the method of tensor analysis. Equations (A.6) can be expressed in the form[9]

$$
-\epsilon_{ij} = s^E_{ij\alpha\beta} X_{\alpha\beta} + d_{ij\nu} E_\nu
\tag{A.57}
$$

where ϵ_{ij} is the second rank strain tensor, $X_{\alpha\beta}$ the second rank stress tensor, $s_{ij\alpha\beta}$ the fourth rank compliance tensor, E_ν the field vector, and $d_{ij\nu}$ the third rank piezoelectric tensor. By employing the geometric rules for tensor

[7] This method of determining the constants for rotated axes is discussed in a former paper "Dynamic Measurements of the Constants of Rochelle Salt," *Phys. Rev.*, April 15, 1939, Appendix I.

[8] This method is discussed in a recent paper by W. L. Bond, "The Mathematics of The Physical Properties of Crystals," *B. S. T. J.*, Jan. 1943.

[9] The tensor method of writing the elastic and piezoelectric relations is discussed by Atanasoff and Hart and by Lawson. See references (2) and (3).

transformation of axes, the components of the rotated tensors are easily calculated and the elastic and piezoelectric constants for rotated crystals determined.

The variation of Young's modulus as a function of orientation was first worked out by Voigt. In terms of the *IRE* angles specifying the orientation of a crystal plate, the *s* compliance modulus (inverse of Young's Modulus) is given by the equation

$$s_{11}^{E'} = s_{11}^{E}(\cos^2 \theta \cos^2 \psi + \sin^2 \psi)^2 + (2s_{13} + s_{44}^{E}) \sin^2 \theta \cos^2 \psi$$

$$\times (\cos^2 \theta \cos^2 \psi + \sin^2 \psi) + s_{33} \sin^4 \theta \cos^4 \psi - 2s_{14}^{E} \sin \theta \sin \psi \cos \psi \quad \text{(A.58)}$$

$$\times [3(\cos \varphi \cos \theta \cos \psi - \sin \varphi \sin \psi)^2 - (\sin \varphi \cos \theta \cos \psi - \cos \varphi \sin \psi)^2]$$

As discussed in Chapter II by W. L. Bond,[10] the *IRE* angles are measured as follows: Taking the X' axis along the length of the crystal, the Y' along the width, and the Z' along the thickness, the angle θ is the angle between the Z or optic axis and Z'. φ is the angle between the projection of the Z' axis on the XY plane and the X axis, while ψ the skew angle is the angle between the length and the tangent to the great circle which contains the Z and Z' axes and the length of the crystal X'. A crystal having its thickness along the X axis (X-cut crystal) will have the angles

$\theta = 90°$; $\varphi = 0°$; ψ variable but equal to 90° when the length coincides with the Y axis. Under these conditions

$$s_{11}^{E'} = s_{11}^{E} \sin^4 \psi + (2s_{13} + s_{44}^{E}) \sin^2 \psi \cos^2 \psi$$
$$+ s_{33} \cos^4 \psi - 2s_{14}^{E} \sin^3 \psi \cos \psi \quad \text{(A.59)}$$

This equation has been made use of in evaluating the elastic constants of quartz as shown by equations (A.20). For this equation A_2 was measured from the Y axis rather than from the Z as in the *IRE* angle and

$$A_2 = \psi - 90° \quad \text{(A.60)}$$

Since from equation (A.19) the frequency of a long thin crystal in longitudinal motion is known to be

$$f = \frac{1}{2\ell} \sqrt{\frac{1}{\rho s_{11}^{E'}}} \quad \text{(A.19)}$$

the longitudinal frequency of any oriented crystal can be calculated from equations (A.58) and (A.19).

It is the purpose of this section to show also that the temperature coefficient of the longitudinal frequency of any oriented crystal can be calculated provided we know the temperature coefficient of the six elastic constants of

[10] Methods for Specifying Quartz Crystal Orientation and their Determination by Optical Means," this issue of the *B. S. T. J.*

quartz, and that regions of low temperature coefficient crystals can be located for this and other simple modes of motion for which the frequency can be calculated in terms of the elastic constants.

Differentiating equation (A.19) with respect to t the temperature

$$\frac{df}{dt} = -\frac{1}{2\ell}\sqrt{\frac{1}{\rho s_{11}^{E'}}}\left[+\frac{d\ell}{dt}\frac{dt}{\ell}+\frac{1}{2}\left[\frac{\frac{d\rho}{dt}}{\rho}+\frac{\frac{ds_{11}^{E'}}{dt}}{s_{11}^{E'}}\right]\right] \quad \text{or}$$

(A.61)

$$\frac{\frac{df}{dt}}{f} = T_f = -T_\ell - \tfrac{1}{2}[T_\ell + T_{s_{11}^{E'}}]$$

where T_α the temperature coefficient of the quantity α is defined as the rate of change of α with temperature divided by the value of α. The temperature coefficient of the length $\ell = X'$ is 7.8 parts per million per degree centigrade along the optic axis, and 14.3 parts per million perpendicular to it. For a general orientation, the temperature coefficient of length varies as

$$T_\ell = 14.3 - 6.5(\sin^2\theta\cos^2\psi) \tag{A.62}$$

Since the total mass remains the same when the crystal expands, the temperature coefficient of the density is the negative of the sum of the coefficients of the three axes or

$$T_\rho = -36.4 \tag{A.63}$$

Hence the temperature coefficient of frequency becomes

$$T_f = 3.9 + 6.5\sin^2\theta\cos^2\psi - \frac{1}{2}\left(\frac{\frac{ds_{11}^{E'}}{dt}}{s_{11}^{E'}}\right). \tag{A.64}$$

Differentiating equation (A.58) we have as the temperature coefficient of a general orientation

$$T_f = 3.9 + 6.5\sin^2\theta\cos^2\psi$$

$$-\frac{1}{2}\frac{\left[\begin{array}{l} s_{11}^E T_{s_{11}^E}(\cos^2\theta\cos^2\psi + \sin^2\psi)^2 + (2s_{13}T_{s_{13}} + s_{44}^E T_{s_{44}^E}) \times \\ \qquad\qquad \sin^2\theta\cos^2\psi(\cos^2\theta\cos^2\psi + \sin^2\psi) + s_{33}T_{s_{33}} \times \\ \sin^4\theta\cos^4\psi - 2s_{14}^E T_{s_{14}^E}\sin\theta\sin\psi\cos\psi \times \\ [3(\cos\varphi\cos\theta\cos\psi - \sin\varphi\sin\psi)^2 - (\sin\varphi\cos\theta\cos\psi + \cos\varphi\sin\psi)^2] \end{array}\right]}{\left[\begin{array}{l} s_{11}^E(\cos^2\theta\cos^2\psi + \sin^2\psi)^2 + (2s_{13} + s_{44}^E)\sin^2\theta\cos^2\psi \times \\ \qquad\qquad (\cos^2\theta\cos^2\psi + \sin^2\psi) + s_{33}\sin^4\theta\cos^4\psi \\ -2s_{14}^E\sin\theta\sin\psi\cos\psi[3(\cos\varphi\cos\theta\cos\psi - \sin\varphi\sin\psi)^2 \\ \qquad\qquad - (\sin\varphi\cos\theta\cos\psi + \cos\varphi\sin\psi)^2] \end{array}\right]} \tag{A.65}$$

Hence since the elastic constants are definitely known, the temperature coefficient of any longitudinally vibrating crystal can be obtained when the separate temperature coefficients are evaluated.

The temperature coefficients appearing in equation (A.65) can all be evaluated from the temperature coefficient angle curves for X cut rotated crystals shown by Fig. 1.19. For an X cut crystal equation (A.65) reduces to

$$T_f = 3.9 + 6.5 \cos^2 \psi$$
$$-\frac{1}{2} \left[\frac{s_{11}^E T_{s_{11}^E} \sin^4 \psi + (2s_{13} T_{s_{13}} + s_{44}^E T_{s_{44}^E}) \sin^2 \psi \cos^2 \psi + s_{33} T_{s_{33}} \cos^4 \psi - 2s_{14}^E T_{s_{14}^E} \sin^3 \psi \cos \psi}{s_{11}^E \sin^4 \psi + (2s_{13} + s_{44}^E) \sin^2 \psi \cos^2 \psi + s_{33} \cos^4 \psi - 2s_{14}^E \sin^3 \psi \cos \psi} \right] \quad \text{(A.66)}$$

The value of $T_{s_{11}^E}$ is obtained directly for $A_2 = 0$ or $\psi = 90°$, for $T_f = -2$ and hence

$$T_{s_{11}^E} = 11.8 \quad \text{(A.67)}$$

Taking three other angles and solving for the remaining constants we find

$$T_{s_{14}^E} s_{14}^E = -5310; \quad (2s_{13} T_{s_{13}} + s_{44}^E T_{s_{44}^E}) = 45{,}130;$$
$$T_{s_{33}} s_{33} = 17{,}400. \quad \text{(A.68)}$$

Inserting the values found for the elastic constants, two temperature coefficients are determined, and one relation is given between the others,

$$T_{s_{14}^E} = +119; \quad T_{s_{33}} = 182; \quad T_{s_{44}^E} - .1112 \, T_{s_{13}^E} = 228.2 \quad \text{(A.69)}$$

The values of (A.68) are sufficient to determine the temperature coefficient of long thin crystals cut at any angle, for inserting these values in (A.65) the temperature coefficient for any oriented crystal in longitudinal vibration is given by

$$T_f = 3.9 + 6.5 \sin^2 \theta \cos^2 \psi$$
$$- \left[\frac{+755 (\cos^2 \theta \cos^2 \psi + \sin^2 \psi)^2 + 22{,}565 \sin^2 \theta \cos^2 \psi (\cos^2 \theta \cos^2 \psi + \sin^2 \psi) + 8700 \sin^4 \theta \cos^4 \psi + 5310 \sin \theta \sin \psi \cos \psi [3(\cos \varphi \cos \theta \cos \psi - \sin \varphi \sin \psi)^2 - (\sin \varphi \cos \theta \cos \psi + \cos \varphi \sin \psi)^2]}{127.9 (\cos^2 \theta \cos^2 \psi + \sin^2 \psi)^2 + 175.8 \sin^2 \theta \cos^2 \psi (\cos^2 \theta \cos^2 \psi + \sin^2 \psi) + 95.6 \sin^4 \theta \cos^4 \psi + 89.2 \sin \theta \sin \psi \cos \psi [3(\cos \varphi \cos \theta \cos \psi - \sin \varphi \sin \psi)^2 - (\sin \varphi \cos \theta \cos \psi + \cos \varphi \sin \psi)^2]} \right] \quad \text{(A.70)}$$

The only regions of low temperature coefficients are the regions for which the two big middle terms are small which requires that $\theta \to 0$, or $\psi \to 90°$.

The first region would be a Z-cut crystal with its length somewhere in the XY plane and would result in a temperature coefficient of two parts per million negative. Such a crystal is not of much interest since there is no piezoelectric constant for driving it. The other region $\psi \to 90°$ would also result in the length being near the XY crystallographic plane, but would allow the major surface to be made perpendicular to the X axis and hence would allow the crystal to be driven piezoelectrically. By allowing ψ to be slightly greater than 90°, the fourth term in the numerator can be made slightly negative and of a value greater than the two positive terms. This results in the +5° X-cut crystal having nearly a zero coefficient and this angle is the most favorable one for a low coefficient longitudinal mode of motion. All other directions have a negative temperature coefficient.

The remaining temperature coefficients of the six elastic constants can be evaluated from Fig. 1.12, and equation (A.22). The frequency temperature coefficient can be expressed by the equation:

$$T_f = 3.9 + 6.5 \cos^2 \theta$$
$$+ \frac{1}{2} \left[\frac{c_{66}^E T_{c_{66}^E} \sin^2 \theta + c_{44}^E T_{c_{44}^E} \cos^2 \theta + T_{c_{14}^E} c_{14}^E \sin 2\theta}{c_{66}^E \sin^2 \theta + c_{44}^E \cos^2 \theta + c_{14}^E \sin 2\theta} \right] \quad \text{(A.71)}$$

since in terms of the *IRE* angles the series of crystals is given by $\varphi = -90$; $\theta = 90 - A_2$; $\psi = 90°$. Taking the AT, BT, and Y-cut, whose coefficients have accurately been determined, we have

<div align="center">TABLE VI</div>

Crystal Cut	Value of A_2	Value of θ	T_f	$T_{c_{66}}^{E'}$	$c_{66}^{E'}$
AT	+35.25°	54.75°	0	−12.0	29.39×10^{10} dynes/cm²
Y	0	90	+86	164.2	40.50
BT	−49	139 or −41	0	−15.2	68.86

From these data and equation (A.71), the three temperature coefficients can be evaluated as

$$T_{c_{66}^E} = 164.2; \qquad T_{c_{44}^E} = 165.7; \qquad T_{c_{14}^E} = +90.2 \quad \text{(A.72)}$$

To convert these into compliance temperature coefficients we have to make use of the relations of equations (A.8)

$$s_{66}^E = 2(s_{11}^E - s_{12}^E) = \frac{c_{44}^E}{c_{44}^E c_{66}^E - c_{14}^{E2}}; \qquad s_{14}^E = \frac{-c_{14}^E}{2(c_{44}^E c_{66}^E - c_{14}^{E2})};$$

$$s_{44}^E = \frac{c_{66}^E}{c_{44}^E c_{66}^E - c_{14}^{E2}}.$$

Differentiating these with respect to t, we have

$$T_{s_{66}^E} = T_{c_{44}^E} - \left[\frac{c_{44}^E c_{66}^E}{c_{44}^E c_{66}^E - c_{14}^{E2}}(T_{c_{44}^E} + T_{c_{66}^E})\right] + \frac{2c_{14}^{E2}}{c_{44}^E c_{66}^E - c_{14}^{E2}} T_{c_{14}^E}$$

$$= \frac{s_{11}^E}{s_{11}^E - s_{12}^E} T_{s_{11}^E} - \frac{s_{12}^E}{s_{11}^E - s_{12}^E} T_{s_{12}^E}$$

$$T_{s_{14}^E} = T_{c_{14}^E} - \left[\frac{c_{44}^E c_{66}^E}{c_{44}^E c_{66}^E - c_{14}^{E2}}(T_{c_{44}^E} + T_{c_{66}^E})\right] + \frac{2c_{14}^{E2}}{c_{44}^E c_{66}^E - c_{14}^{E2}} T_{c_{14}^E}$$

$$T_{s_{44}^E} = T_{c_{66}^E} - \left[\frac{c_{44}^E c_{66}^E}{c_{44}^E c_{66}^E - c_{14}^{E2}}(T_{c_{44}^E} + T_{c_{66}^E})\right] + \frac{2c_{14}^{E2}}{c_{44}^E c_{66}^E - c_{14}^{E2}} T_{c_{14}^E}$$

(A.73)

TABLE VII

Temperature Coefficient	Present Determination	Previous Determination	Bechmann
$T^E_{s_{11}}$	+11.8	+12	+11.5
$T^E_{s_{12}}$	−1352	−1265	−1125
$T_{s_{13}}$	−294.8	−238	−148
$T^E_{s_{14}}$	+120	+123	+113
$T_{s_{33}}$	+182	+213	+180
$T^E_{s_{44}}$	+195.4	+189	+175
$T^E_{s_{66}}$	−134.2	−133.5	−119

TABLE VIII

Temperature Coefficient	Present Determination	Previous Determination	Atanasoff & Hart	Bechmann	Koga
$T^E_{c_{11}}$	−46.5	−54	−49.7	−48	−61.1
$T^E_{c_{12}}$	−3300	−2350	−3000	−2115	—
$T_{c_{13}}$	−697	−687	−580	−530	—
$T^E_{c_{14}}$	+90.2	+96	+107	+82	+110
$T_{c_{33}}$	−204.5	−251	−213	−208	—
$T^E_{c_{44}}$	−165.7	−160	−169	−151	−199
$T^E_{c_{66}}$	+164.2	+161	+170.1	+144	+199

Inserting the numerical values for the elastic constants and the temperature coefficients we have

$$T_{s_{66}^E} = .883T_{s_{11}^E} + .1071T_{s_{12}^E} = -134.5; \qquad T_{s_{14}^E} = 121.4;$$

$$T_{s_{44}^E} = 195.4$$

(A.74)

The value of $T_{s_{14}^E}$ provides a check on the accuracy of measurement since it has been measured in two ways. The agreement is within about 2 per cent which shows the probable accuracy of the measurement. Combining the coefficients of (A.69) with those given by equation (A.74), the complete temperature coefficients are given in Table VII together with previous determinations[11,12]. The present determination differs from a previous determination by the writer due to the use of the elastic constants found here rather than Voigt's constants.

The temperature coefficients of the c_{ij}^E elastic constants are easily obtained from the s_{ij}^E constants by employing the relations of equation (A.8). These result in the temperature coefficient values for the c constants given in Table VIII.

By using the elastic constant data, the temperature coefficient data, and the equations of transformation for rotated axes it is possible to calculate the frequency and temperature coefficient of any simple mode for any orientation. Examples are given for a face shear mode and a thickness shear mode in a previous paper "Low Temperature Coefficient Quartz Crystals."[13] This paper shows contour maps for low temperature coefficient crystals of these types.

[11] The first determination of the temperature coefficients of the writer was given in a paper "Electrical Wave Filters Employing Quartz Crystals As Elements," *B. S. T. J.*, July 1934, p. 446. A redetermination using better temperature coefficient data was given in a paper "Low Temperature Coefficient Quartz Crystals," *B. S. T. J.*, Jan. 1940. The present determination uses the same temperature coefficient data but slightly different elastic constants which results in slight changes in the temperature coefficients.

[12] A partial determination of the coefficients was made by Koga, Rep. Rad. Research, Japan 6, 1 (1934). Other complete determinations are R. Bechmann, Hoch: *tech. U. Elek. Akus.* 44,145 (1934) and Atanasoff and Hart, *Phys. Rev.*, Vol. 59, No. 1, Jan. 1, 1941, pp. 85, 96.

[13] *B. S. T. J.*, Vol. XIX, 74 (Jan. 1940).

Methods For Specifying Quartz Crystal Orientation and Their Determination by Optical Means

By W. L. BOND

2.1 QUARTZ AND ITS AXES

The chemist describes quartz as silicon dioxide, SiO_2, crystallized in hard, brittle, glass-like, six sided prisms, often with pyramidal terminations; melting point 1750° Centigrade, density 2.65, hardness on Moh's scale 7. It transforms from alpha to beta quartz at 573°C under atmospheric pressure. Under stress it transforms at lower temperatures. Alpha quartz is insoluble in ordinary acids but soluble in hydrofluoric acid; and in hot alkalis.

At first glance we might say that it had hexagonal symmetry but if we etch two adjacent pyramid faces we find that the microscopic etch pits are of different shape, hence the faces cannot be equivalent. It has three axes of two-fold symmetry and one axis of three-fold symmetry. Let us also remark that it does *not* have a center of symmetry or a six-fold axis. Figure 2.1 shows us that the three two-fold axes are perpendicular to the three-fold axis and are 120° apart. If they were not like this, they would not be self-consistent.

As we examine more and more quartz crystals we find that there is a tendency for pyramid faces to be alternately large and small, the larger faces being brighter than the smaller faces. Also the etch pits of *alternate* faces are similar. (The etch pit study is a powerful tool in determining crystal symmetry.) Further, two other "kinds" of faces are quite commonly found. If we draw such a crystal as though equivalent faces were of equal size we get such a picture as Fig. 2.2. It is an idealized figure used to illustrate the symmetry of quartz. The prism faces are marked *m*, the six faces marked *r* "constitute the primary rhombohedron"—the ones we called the large bright pyramid faces. The crystallographer thinks of these six faces as pieces of the faces of a rhombohedron. (A crystallographer's rhombohedron is like a cube stood on one corner, then the opposite corner pushed in a little towards the other, or pulled away from it. He thinks of it always as standing on this corner, as Fig. 2.3.) The *z* faces constitute a second rhombohedron—the secondary rhombohedron or minor pyramid faces. The *s* and *x* faces illustrate a further property of quartz. Figure 2.3 differs from its mirror image so that we

might expect to find two kinds of quartz that are related to each other as one's right hand is related to his left. We do find them and call them right-hand quartz and left-hand quartz respectively. They are illustrated in Fig. 2.4. These conventional figures are shown in many texts but no one has seen such perfect quartz crystals. They are drawn possessing just these faces and no others merely to illustrate the symmetry of quartz and its occurrence in right-handed and left-handed forms.

These figures are also useful in defining how a blank shall be cut from one kind of quartz. It is found that if a crystal be compressed with forces
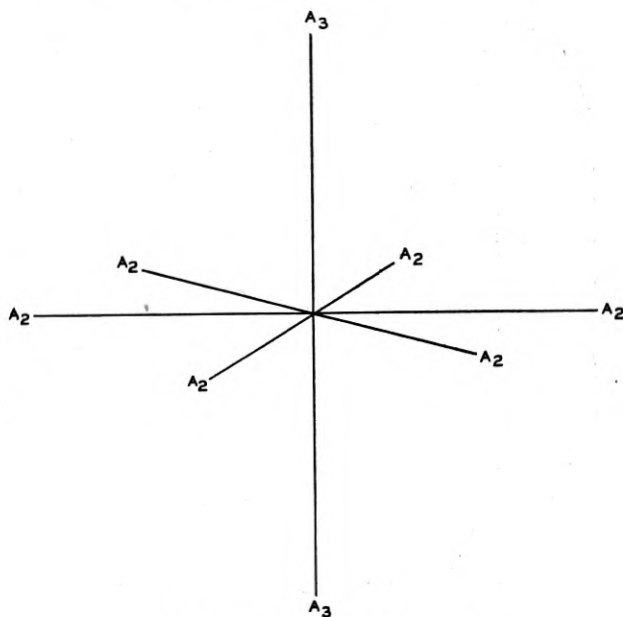
Fig. 2.1—Hexagonal axis system

parallel to a pair of sides of the hexagon an electric polarization takes place in the direction of the forces. The edge "modified" by the presence of s and x faces becomes negative. If we allow these charges to leak off and then suddenly release the mechanical forces the "modified" edge becomes positively charged as the crystal expands. This is true for both right-hand and left-hand crystals.

Let us now talk about right-hand quartz. Since expansion is considered as a positive strain (contraction as negative) it is decided to take the positive end of the electric axis as pointing towards the modified edge. This gives us a positive charge at the positive end of the electric axis when a positive stress (tension) is applied along this axis. This positive direc-

tion of an electric axis is taken as the positive x axis of a right-hand xyz rectangular coordinate system. The z axis is taken along the axis of the hexagonal prism, and since the x axis is an axis of two-fold symmetry we
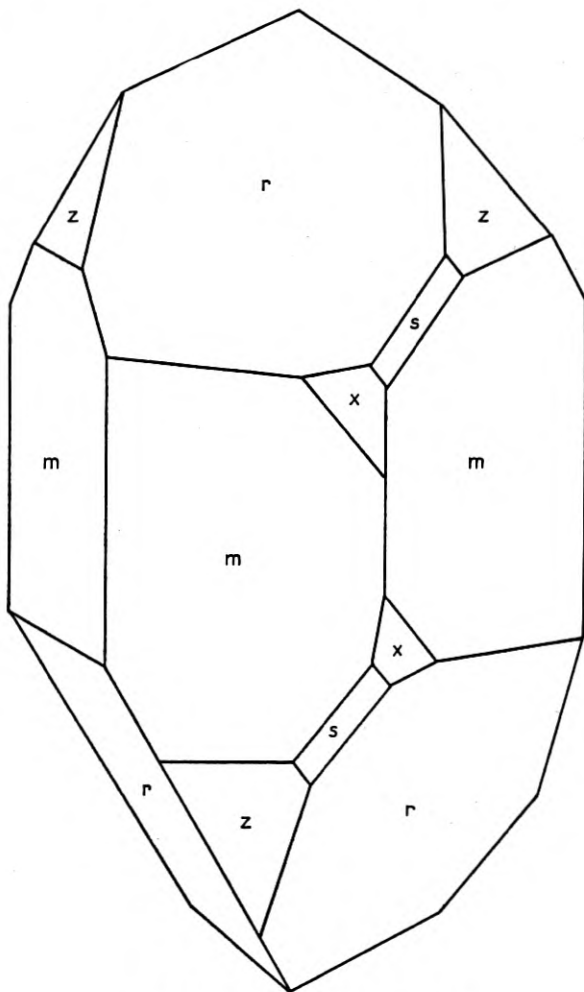
Fig. 2.2—An idealized quartz crystal

can take either end of the prism as the direction of $+z$. We then choose y to form a right-handed coordinate system. (In a right-handed system if a right-handed screw turns about the z axis in the sense x to y it would ad-

vance in the positive z direction.)   The y axis will always lie directly under a major rhombohedral face.

We could define the x, y and z axes for a left-hand crystal as forming a left-hand system.   Though this is a useful conception in mathematical

Fig. 2.3—A rhombohedron
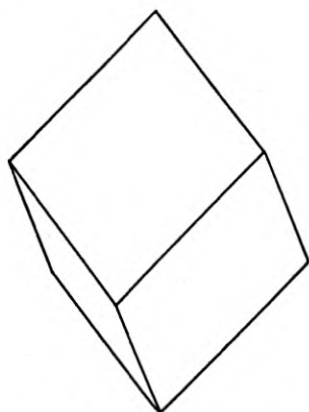
LEFT HAND
QUARTZ
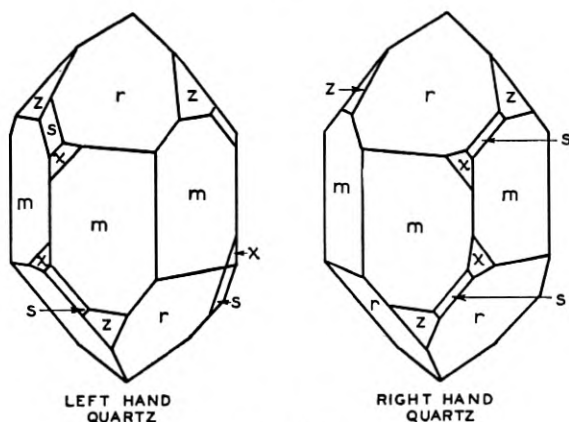
RIGHT HAND
QUARTZ

Fig. 2.4

studies, we can dodge this "double standard" by a simple device.   For use as a crystal circuit element, left-hand quartz can be used just as well as right-hand quartz.   In designing such an element it suffices to think always in terms of right-hand quartz and issue specifications for this kind only, using always right-hand coordinate systems.   If now for left-hand
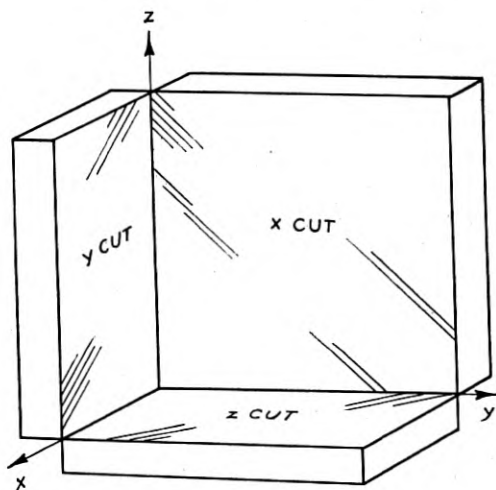
Fig. 2.5—Simple crystal cuts
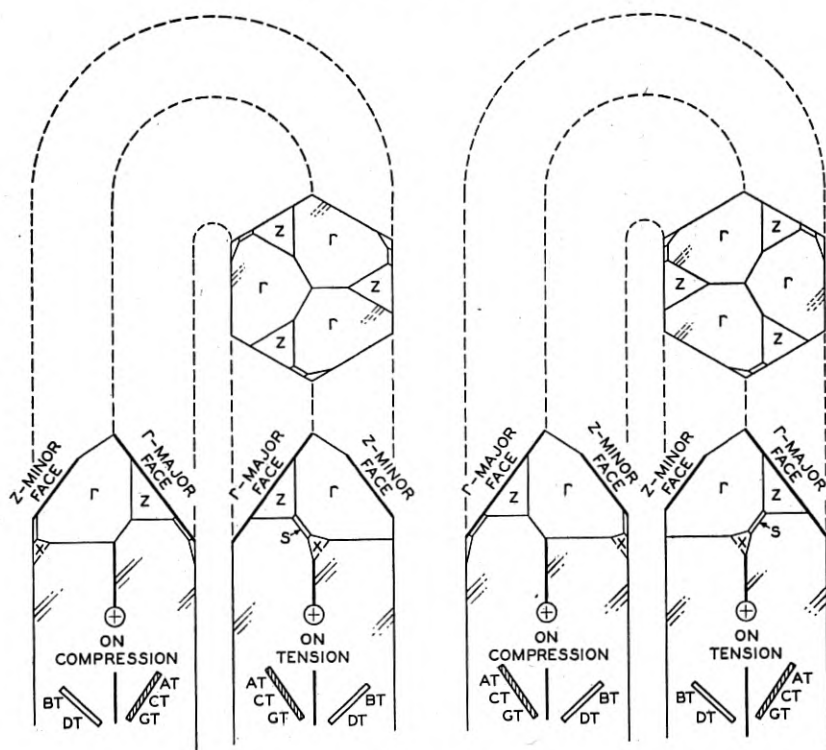
Fig. 2.6—Simply rotated cuts

crystals we mark the negative end of the electric axis as positive we can treat it exactly as though it were a right-handed crystal.

The first plates used were x and y piezoids (squeezing solids). For these simply described cuts one does not need to know the quartz "handedness." These crystals had large frequency temperature coefficients. But when Lack, Willard and Fair brought out the low temperature coefficient AT plate, its more complicated orientation required the right-left differentiation. The AT, the subsequent BT, CT, DT, etc., were thought of as y-cuts rotated through various angles about the edge that coincided with x. For example, the AT was a $+35\frac{1}{4}°$ cut, or was a y plate rotated $35\frac{1}{4}°$ about x; the BT was a $-49°$ cut. Their orientations are illustrated in Fig. 2.6.

As more complicated orientations were designed to give even better temperature coefficients at extreme frequencies the description became more difficult, requiring the specification of two or three angles. Many schemes for specification have been devised but the Institute of Radio Engineers is recommending the adoption of a system we shall call the I.R.E. system.

The crystal designer has the problem: "How shall I orient the length, width and thickness of a piezoid with respect to the x, y and z axes so as to give the required electrical properties? He thinks in terms of fixed x, y and z axes, variable piezoid edge directions. The crystal cutter has the converse problem. "How shall I orient the x, y and z axes of the crystal so that fixed saws will give the required surfaces?" For this reason the most convenient orientation angles from the designer's viewpoint may not be the simplest from the cutter's viewpoint. Also the translation from one set to the other may not be simple.

The early methods of orientation specification were somewhat chaotic. There was no overall plan of what angles were to be specified and from what axes they were to be measured. Each group of crystals was a law unto itself.

The I.R.E. Orientation Angles

The relations between the x, y and z axes of the crystal and the length, width and thickness of the piezoid are given in Fig. 2.7.

The position of Fig. 2.7 may be considered as a result of turning the piezoid through the successive angles ϕ, θ, ψ starting from an initial position length parallel to x, width parallel to y and thickness parallel to z as in Fig. 2.8. First the crystal is turned through angle ϕ about z in the direction shown in Fig. 2.7. Then it is lowered through angle θ about an axis parallel to the width direction, again in the direction shown in Fig. 2.7. Finally it is skewed through an angle ψ about an axis parallel with the thickness direction in the sense shown in Fig. 2.7.
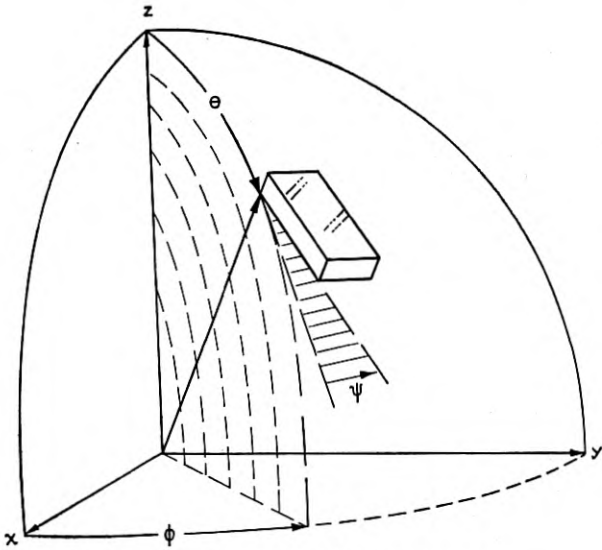
Fig. 2.7—The I.R.E. orientation angles
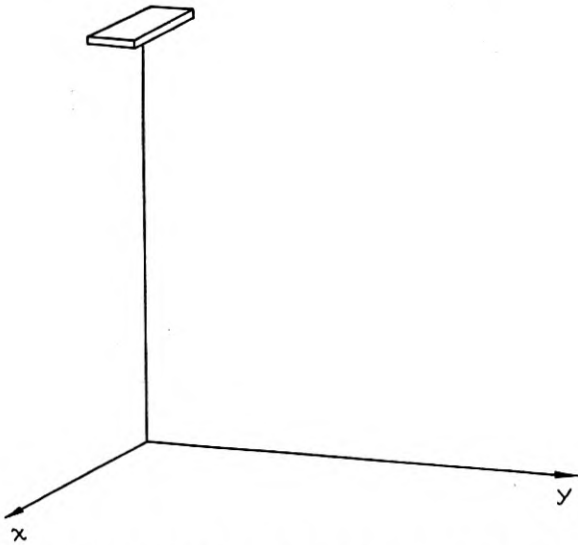
Fig. 2.8—The initial position 0, 0, 0 for the I.R.E. angles

The I.R.E. Angles for a Few Standard Piezoids

Name	ϕ	θ	ψ
z	0	0	0
x	0	90	90
$-18°$ Filter	0	90	108
$+5°$ "	0	90	85
y	90	90	90
AT	-90	$54\frac{3}{4}$	90
BT	-90	-41	90
CT	-90	52	90
GT	-90	$38°\ 52'$	$\pm45°$
MT	$6°\ 40'$	$50°\ 28'$	$79°\ 36'$
NT	$9°\ 25'$	$40°\ 40'$	$77°\ 40'$

2.2 Orientation by Natural Faces

With well faced material one can do an accurate job of orienting without X-rays if he knows the faces of quartz thoroughly.

The quartz rhombohedral faces are highly perfect and polished, the major often more so than the minor. With two such faces a device like that illustrated in Fig. 2.9 could be used to give an orientation accurate to a minute or two. An adjustable base, symbolized here as a ball and socket, is adjusted so that the eye centers the lamp filament image on the cross hairs, first for one face, and then, turning the base about on the reference table it is adjusted for another face. When the images all pass through center as the base is turned on the table the optic axis is perpendicular to the table. When any one image is centered, the electric axis is perpendicular to the plane of the paper. This with the already mentioned fact that AT plates are cut near a minor face and BT's near a major, allows us to cut the crystal accurately.

Although the rhombohedral faces are highly perfect the prism faces never are. On the prism face, true prism faces that are very short in the z direction alternate with short rhombohedral faces to give the general contour a slant. These "steps" give the face a striped look. The stripes are known as growth lines or striations. They are parallel to x and can be used to find x to a degree or so. If we sight on striations on two adjacent faces we can locate the optic axis to nearly the same accuracy since the optic axis is perpendicular to the striations on all faces.

There are several indications that help us find, from the prism, where the major rhombohedron would be in the absence of such faces. Some crystals grow in the form shown in Fig. 2.10. They are symmetrically doubly terminated and a very narrow prism is found under the major rhombohedron, a wide face under the minor. Hence given a portion like that enclosed in the dotted line we could deduce the complete orientation.

Some crystals grew out at right angles to a wall and because they grew
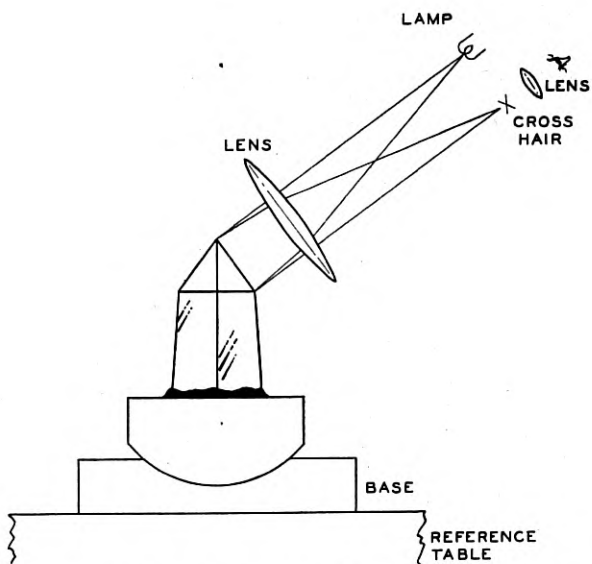
Fig. 2.9—Optical orientation by reflection of light from faces
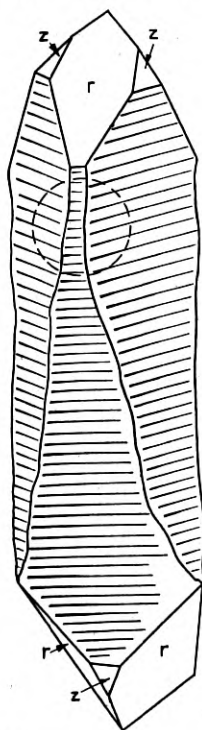
Fig. 2.10—A type of quartz growth

along a z axis in one direction only, the x axis does not appear to be one of two-fold symmetry. Such a crystal is illustrated in Fig. 2.11. Here the prism faces under a major rhombohedron are tapered and bright, the prism faces under a minor are relatively parallel sided and very dull. The bright prism faces are much more nearly parallel to the optic axis than the dull ones. Again, given a portion of the prism we can deduce the orientation.
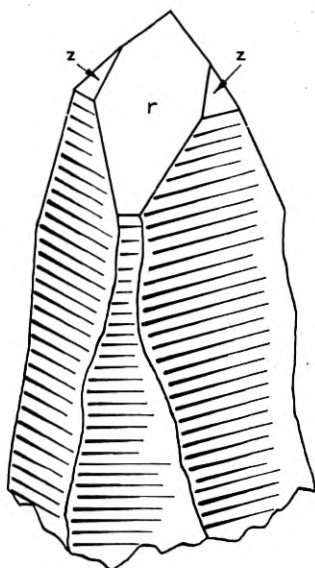
Fig. 2.11—Another type of quartz growth

2.3 FLAT LAY CUTTING

Flat lay cutting takes advantage of the fact that, although tapering quartz prisms have their faces non-parallel to the z axis the prism faces are parallel to the x axis. A crystal is cemented prism face down, to a mounting plate. The mounting plate is tilted and turned on a base plate to render the optic axis parallel to the long edge of the base plate. This is done in a conoscope. Now the edges of the base plate are the x, y, z axes of the crystal.

The crystal can now be cut directly into wafers for dicing into AT's, BT's etc. by mounting on an angle bracket as shown in Fig. 2.13 or cut into X sections from which AT or BT bars shall be made by merely sawing down the length. Again z sections can be cut by cross cutting. Good z sections can be made in this way in the total absence of faces. These sections can
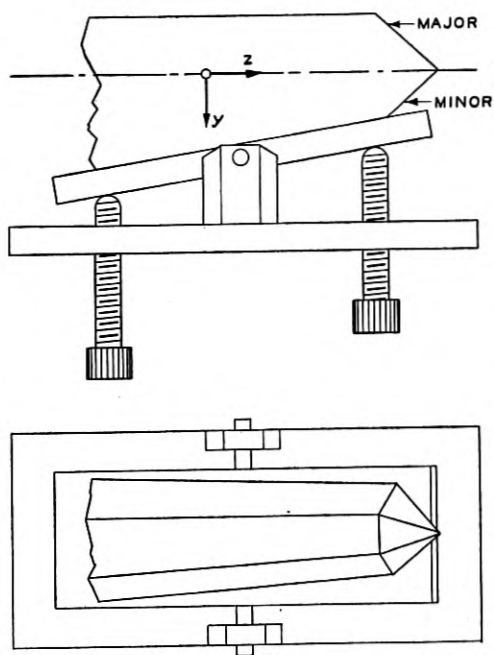
Fig. 2.12—Optical adjustment for the sawing of Z sections or direct crystal blank slabs
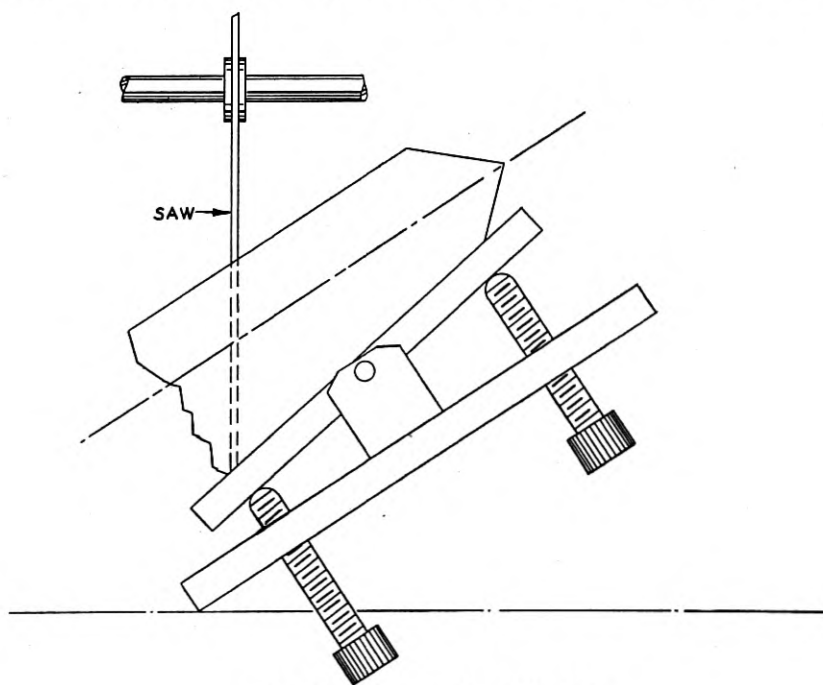
Fig. 2.13—Direct sawing of the slabs

then have their x axes determined by etching and X-rays and cut up by the Z section cutting method.

By turning the base plate on the angle bracket and dicing the wafers at an angle any orientation can be obtained.

2.4 Z SECTION OR VERTICAL CUTTING

Having a true Z basal section it is first marked for the $+x$ axis. The simplest procedure is to use the star mark; for right-hand quartz (R.H.Q.)
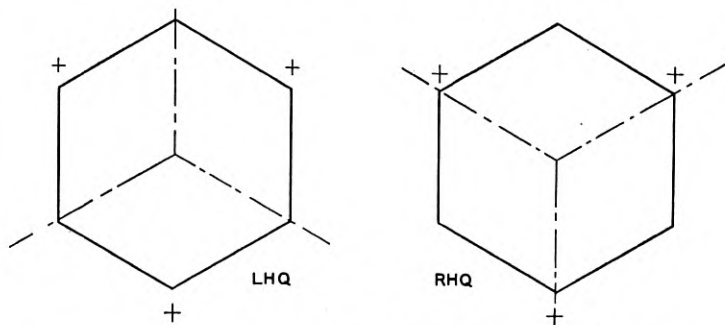
Fig. 2.14—Marking the "sense" of righthand and lefthand sections
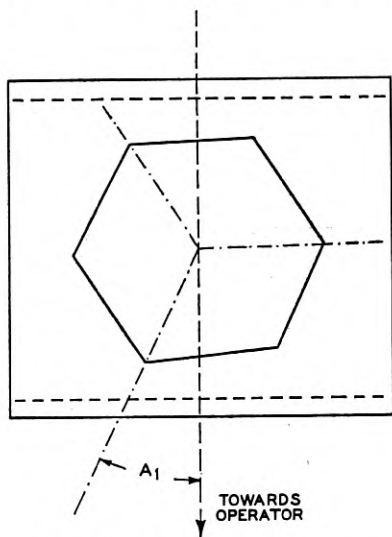
Fig. 2.15—Making the rotations A_1

the rays should point toward the plus electric axis, for left-hand quartz the rays should point towards the negative electric axis.

The section is now placed on the carriage plate, one ray pointing towards the operator (which ray is decided on the basis of the economy of quartz).

SAW
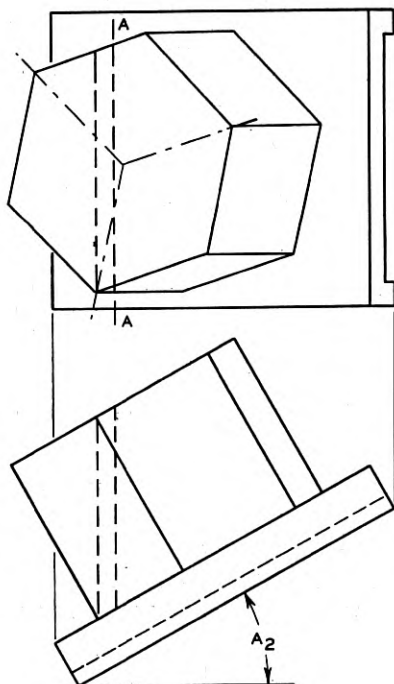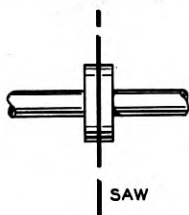
Fig. 2.16—The rotation A_2
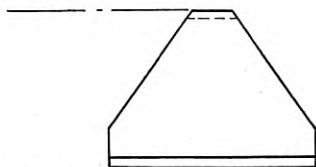
Fig. 2.17—The slab after the rotations A_1 and A_2
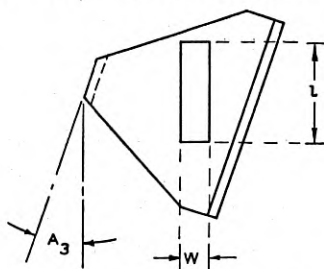
Fig. 2.18—Making the A_3 rotation

236

The section is then rotated clockwise on its base, through angle A_1 as in Fig. 2.15 and cemented in this position.

The carriage plate is then transferred to a diamond saw angle bracket of tilt A_2, as in Fig. 2.16, and the crystal is sawed into slices slightly thicker than the required final thickness t.

The operator turns these slices down flat on the table of a dicing saw as in Fig. 2.17 by rotating the slices 90° clockwise about the axis AA, then turns the slice through angle A_3 as in Fig. 18 and makes a cut. The plate is finished as shown in Fig. 2.18.

Since the angle bracket is not reversible, negative A_2 angles are cut by adding ±180° to A_1 and reversing the sign of A_3.

THE A ANGLES FOR SOME STANDARD PLATES

Cut	A_1	A_2	A_3
x	90°	0	0
y	0	0	0
z	0 or 90°	90°	0 or 90°
-18	90°	0	$+18°$
$+5$	90°	0	-5
AT	0	$35\frac{1}{4}$	0
BT	180°	$+49°$	0
CT	0	38°	0
DT	180°	52°	0
GT	0	51° 7′	$±45°$
MT	96° 40′	39° 32′	$-10°$ 24′
NT	99° 25′	49° 20′	$-12°$ 20′

2.5 THE RELATION BETWEEN THE I.R.E. ANGLES $\Phi\theta\psi$ AND THE Z SECTION ANGLES A_1, A_2, A_3

It can be shown that:

$$A_1 = 90 + \phi$$

$$A_2 = 90 - \theta$$

$$A_3 = -90 + \psi$$

2.6 POLARIZED LIGHT AS APPLIED TO CRYSTALS

Light consists of electromagnetic "vibrations." The vibrations are perpendicular to the direction of propagation but ordinarily helter-skelter in all directions perpendicular to the propagation. The color of the light is determined by the vibration frequency, blue vibrating more rapidly than red. In a vacuum, light travels at 186,000 miles per second (3×10^{10} cms per second) all colors at the same velocity. On entering a transparent medium the velocity is reduced, ordinarily blue being slowed more than red. The frequencies are unaltered on entering the medium.

Light traveling through a uniaxial crystal in the direction of Fig. 2.19 breaks up into two components that travel at different velocities. For one

of these components the vibration is all in the plane of *poz*, of the other the vibration is all perpendicular to the plane of *poz*.
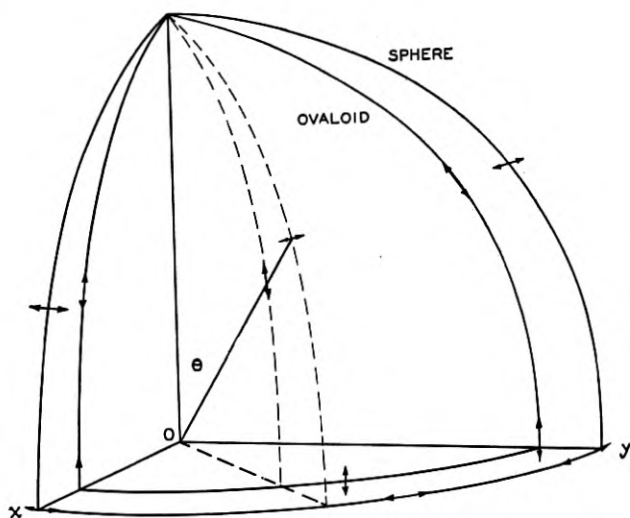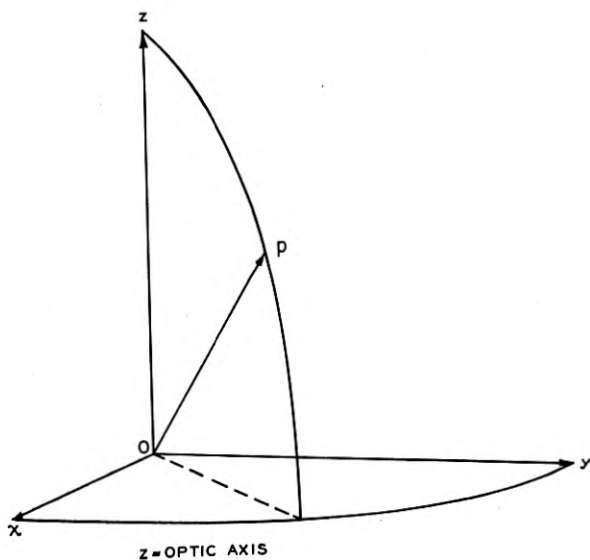
Fig. 2.19—The velocities of light in a uniaxial crystal

A plot of the propagation velocities for all directions is a surface of two sheets, one octant of which is shown in Fig. 2.19. One sheet is a sphere;

the other sheet, which is an ovaloid of revolution, touches the sphere at the two points where the double sheet is pierced by the optic axis. Of the two rays traveling along one line, the one with a velocity corresponding to the sphere is called the ordinary ray; the one with a velocity corresponding to the ovaloid is called the extraordinary ray. For quartz the ovaloid is prolate and lies inside the sphere. For tourmaline the ovaloid is oblate and lies outside the sphere. The small arrows show the direction of vibration. Each of the components is said to be polarized since for each all the vibration is in one direction.

Since both sheets are surfaces of revolution with the optic axis as the axis of revolution, we can never tell the x axis from the y axis by optical means.* Only the z axis is a unique direction and can be determined optically. If this figure is taken to represent the case for blue light there will be a slightly larger but similar figure for red light since, in the crystal, red light travels faster than blue light.

Some kinds of crystals have velocity plots for which the double sheet surfaces touch at four points. Hence they have two optic axes and are called biaxial. All hexagonal, rhombohedral and tetragonal crystals are uniaxial, all others except the isometric ones are biaxial. Rochelle Salt is biaxial.

2.7 POLARIZERS AND ANALYZERS

In the Nicol prism means were found to eliminate the ordinary ray; the other is transmitted as polarized light. That is, ordinary light of any or all colors upon passing through a Nicol prism emerges as plane polarized light with no change in color.

Transparent colored media appear colored because they absorb some colors of light more than other colors. In colored crystals the two rays themselves often differ in their color absorption so that the crystal as viewed by means of the ordinary ray seems to be of a different color than as viewed by the extraordinary ray. Quinine iodo-disulfate, or Herapathite, absorbs most visible colors of one ray almost completely; transmits about 60% of the visible colors in the other ray. Hence light emerging from this crystal is almost completely polarized. In the commercial product called "polaroid", myriads of such crystals, microscopic in size, are contained in a celluloid-like sheet and oriented by stretching the sheet. This material now replaces Nicol prisms for all but the most exacting uses.

If we put two identical polaroid sheets together with their transmission vibration directions parallel as in Fig. 2.20 we can see through them but if their transmission vibration directions are at right angles we cannot see

---

* Methods depending on etch pits are excluded. They are optical only in the sense that observing natural faces is optical.

through them because the second sheet can transmit none of the vibrations transmitted by the first sheet. As we rotate the second sheet back from complete extinction to "best transmission" the transmitted light increases sinusoidally. In any such arrangement as Fig. 2.20 the first sheet is called the polarizer, the second is called the analyzer. The name analyzer is chosen because light that can be extinguished by means of a suitably rotated analyzer must be plane polarized, and it must be vibrating at right angles to the transmission vibration direction of the analyzer when set for extinction. The transmission vibration direction of a polaroid plate will hereafter be called its vibration axis.

Let us go back to Fig. 2.19 and cut out from around the point p, the small tangential crystal plate shown magnified in Fig. 2.21. Here p is the direction of propagation as before, and z is parallel to the optic axis. Also s which is

UNPOLARIZED OR PLAIN LIGHT

POLARIZED LIGHT VIBRATING VERTICALLY
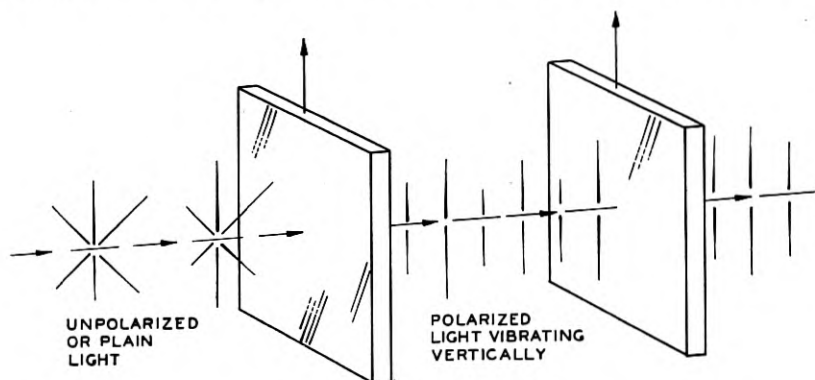
Fig. 2.20—Light polarization

in the plane of p and z, is the direction of slow vibration while f which is perpendicular to this plane is the direction of fast vibration. The vibration frequency is really the same for both. "Slow vibration" means "vibration direction for slow transmission." All directions of propagation that have this vibration axis have the same velocity.

In Fig. 2.22 we have placed this plate between "crossed polaroids"—that is polaroids set for extinction. The slow direction makes an angle α with the polarizer vibration axis. When this vertical-polarized ray of intensity I enters the crystal it breaks up into components, one of intensity:

$$I \cos \alpha \text{ vibrates along s} \tag{2.1}$$

and one of intensity

$$I \sin \alpha \text{ vibrates along } f \tag{2.2}$$

as illustrated in Fig. 2.23.
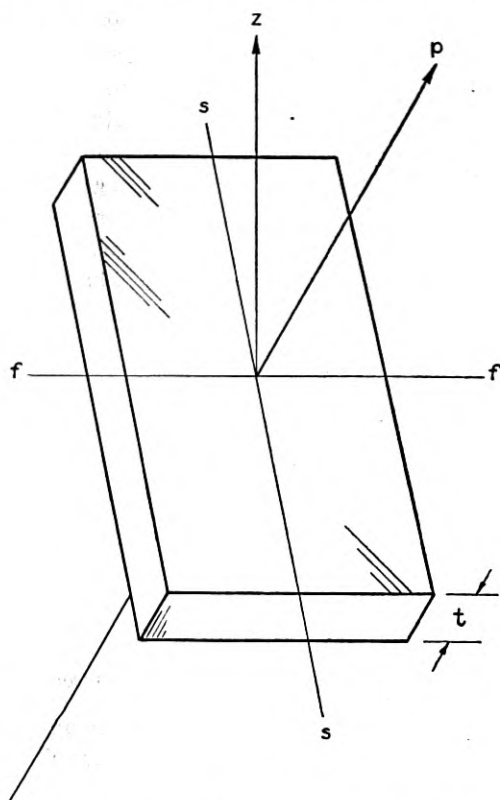
Fig. 2.21—Light breaks into components in the crystal
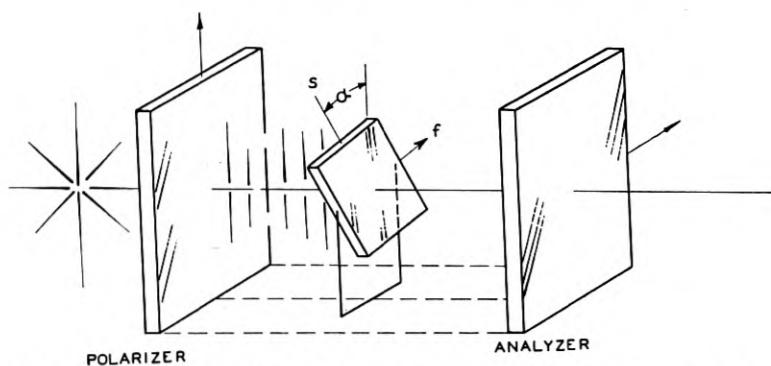
Fig. 2.22—Transmission when a crystal is placed askew between crossed polarizers

If $\alpha = 0$ the fast component reduces to zero and the slow component goes through the crystal unchanged hence emerging plane polarized. It can

then be extinguished by the analyzer. If $\alpha = 90°$ the slow component reduces to zero and the fast one goes through unchanged and again can be extinguished by the analyzer. This effect can be used to check crystal orientations. Such an instrument fitted with a divided circle used to measure α is called a stauroscope. The stauroscope often uses a special analyzer capable of better determination of extinction setting.

If α is not 0 or 90° two components traverse the crystal and recombine at the boundary. These two components are at right angles to each other; they are of unequal intensities, and they differ in phase because they traveled at different speeds.
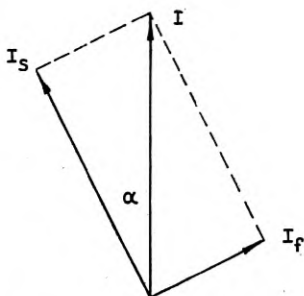
Fig. 2.23—The intensity of the two components from Fig. 22

Now v_s and v_f have the same frequency F so that in unit time each makes F wave-lengths. This requires that the slow ray have F wave-lengths in a distance v_s and hence that each wave have a length:

$$\lambda_s = \frac{v_s}{F} \qquad (2.3)$$

Similarly

$$\lambda_f = \frac{v_f}{F} \qquad (2.3')$$

In a distance t there are $\frac{t}{\lambda_f}$ fast waves and $\frac{t}{\lambda_s}$ slow ones. Let us say that there are N more fast waves than slow ones in the distance t. Consequently $N = \frac{t}{\lambda_f} - \frac{t}{\lambda_s}$ which, from (2.3) and (2.3') may be written:

$$N = \frac{tF}{v_f} - \frac{tF}{v_s} \qquad (2.4)$$

All the data on light are given in terms of wave-lengths in a vacuum, not in terms of frequency, so we will assume that in a vacuum the wave-length

of this light is λ, and as in a vacuum its velocity is $V(= 3 \times 10^{10}$ cms per second) an equation similar to (2.3) would tell us that $\lambda = \dfrac{V}{F}$ and hence that:

$$F = \frac{V}{\lambda} \tag{2.5}$$

With (2.5) we can rewrite (2.4) as

$$N = \frac{t}{\lambda}\left(\frac{V}{v_f} - \frac{V}{v_s}\right) \tag{2.6}$$

The ratio of the velocity in a vacuum to the velocity in a medium is called the refractive index of the medium commonly given the symbol n. For most transparent materials n is between 1.3 and 1.8.

We write these refractive indices as

$$\frac{V}{v_f} = n_f \quad \text{and} \quad \frac{V}{v_s} = n_s \quad \text{respectively.}$$

Now (2.4) becomes:

$$N = \frac{t}{\lambda}(n_f - n_s) \tag{2.7}$$

After passing through the crystal plate of thickness t, Fig. 2.21, the two light components recombine. They are polarized at right angles each to each; they are of unequal intensities, and they differ in phase by N wavelengths as given by equation (2.7).

If the crystal were vanishingly thin the two components that recombine would be effectively in step or in phase. In Fig. 2.24 we have plotted vibration amplitude against time for these two components. They are separated for clarity. In the upper curve the slow vibration is shown as vertical, in the center curve the fast one is shown as horizontal. In the lower curve corresponding points have been added vectorially. From actual construction we see that in the resultant curve the vibration is always parallel to the line AA'. Hence it is plane polarized and can be extinguished by means of an analyzer.

Let us now consider the case of a thicker plate for which the slow wave gets a quarter-wave-length behind the faster one. This case is plotted in Fig. 2.25 in the same way that the previous case was plotted in Fig. 2.24. It turns out to be a space curve like a slightly flattened corkscrew. Viewed along the axis it looks like an ellipse, as shown to the right of the space curve. If the slow ray had lost but a little with respect to the fast ray, we would have gotten a very flat ellipse. If the two components had had the

same amplitude with the quarter wave phase difference the end view in Fig. 2.25 would have been a true circle. Now since the vibration, in these cases, is not all in one plane, the light cannot be extinguished by an analyzer; it is not plane polarized light. In the one case it is called "elliptically polarized" light and in the other, "circularly polarized" light. If the slow ray loses an integral number of wave-lengths, it makes no difference; only frac-
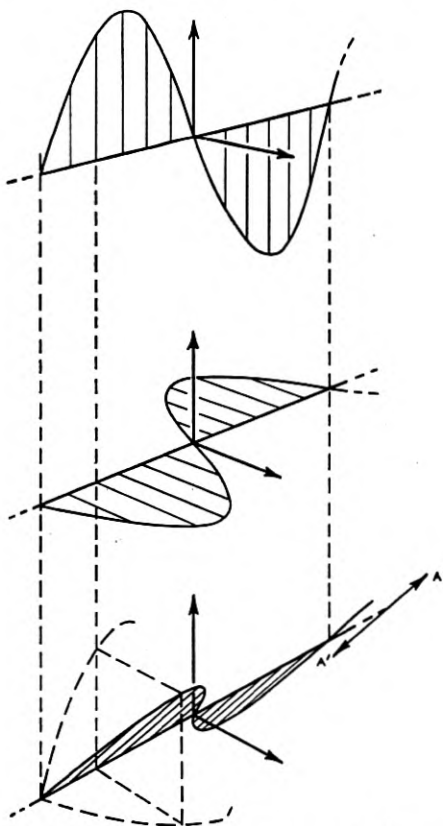
Fig. 2.24—The recombination of the light components after passing through a thin crystal

tions of wave-lengths count, except that if several wave-lengths are present a thickness that is right to give an integral number of wave-lengths for one color may give an integer plus a fraction for some other color. If the thickness is fairly small, this may cause spectral colors from white light. For thick plates the wave-lengths so overlap that the field appears colorless but dark or white according to the value of the angle α; i.e., if α is zero or

90° the field is dark; if $\alpha \neq 0$ or 90° the field is bright.   In Fig. 2.26 we illustrate how, for one color of light the polarization starts out as plane at the
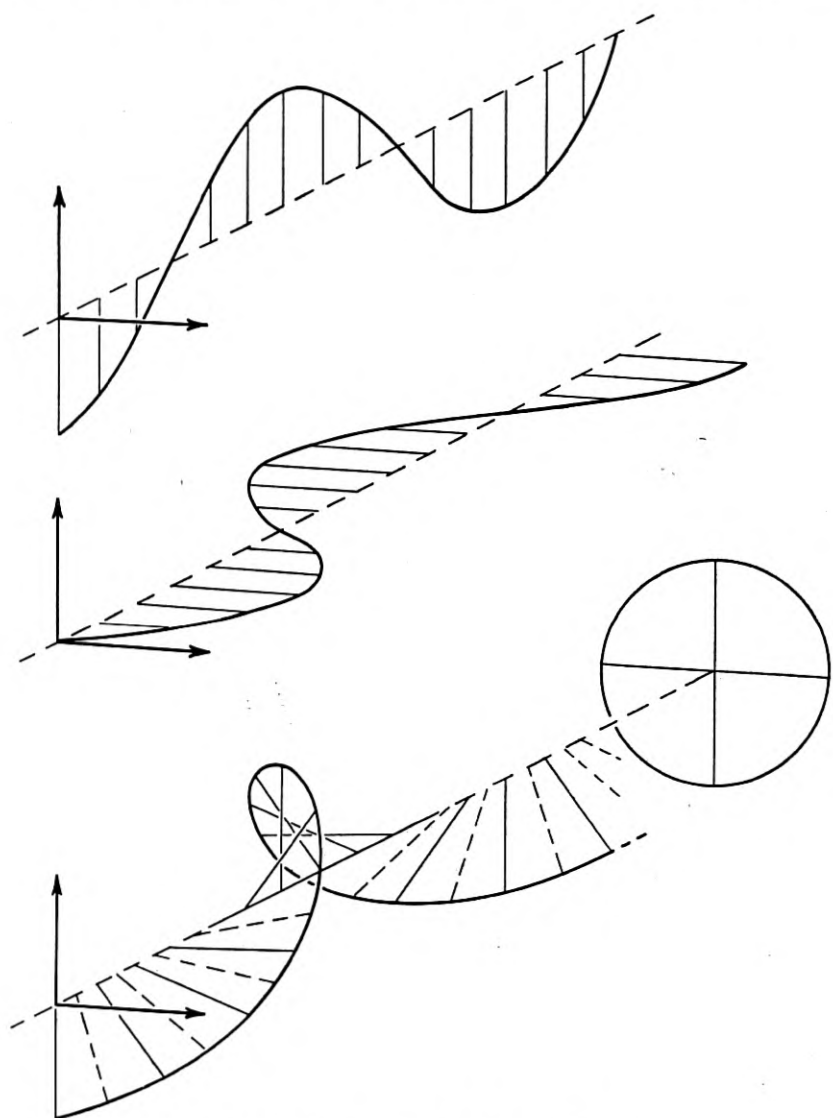
Fig. 2.25—Recombination more generally

crystal boundary, passes through elliptical to circular polarization, then flattens out the other way through elliptical to plane polarization at a distance

in the crystal corresponding to the slow wave being one wave-length behind the fast one.

In Fig. 2.27 we show two AT plates resting on a reference surface. In the first crystal the propagation is perpendicular to x, in the second crystal it is
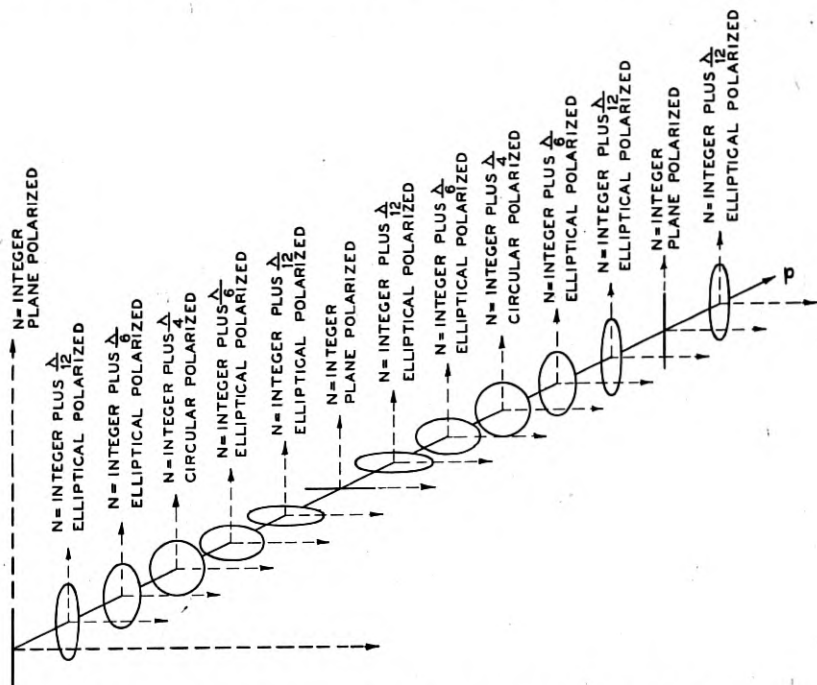
Fig. 2.26—How the kind of polarization changes with crystal thickness
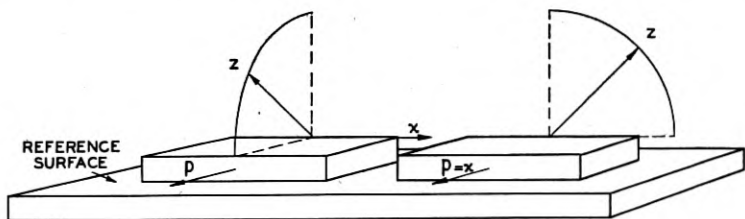
Fig. 2.27—The normascope principle

along x. If the reference surface is the reference table of a simple stauroscope the edge of the first crystal will appear dark because $\alpha = 0°$ or $90°$; the edge of the second crystal will appear bright because α is not $0°$ or $90°$. Actually this α for an AT plate can be $\pm35° \pm$ any multiple of $90°$ because we don't know whether z stands out to the right or to the left. Hence the

reading might be, for instance, $-35°$, $+35°$, $+55°$, $+125°$, etc.   This is the principle of the *normascope* used to identify the x direction for crystal adjustment.

Let us study these relative phase shifts at different angles near the optic axis.   Now quartz has an optical complication beyond that just described— it rotates the plane of polarization of plane polarized light traveling along the optic axis.   This complicates our present attempt to build up a background sufficient for an understanding of the conoscope.   But the cono-scope finds the optic axis for other crystals that do not rotate the plane of polarization, tourmaline for example; so we will ignore this rotation, to be-
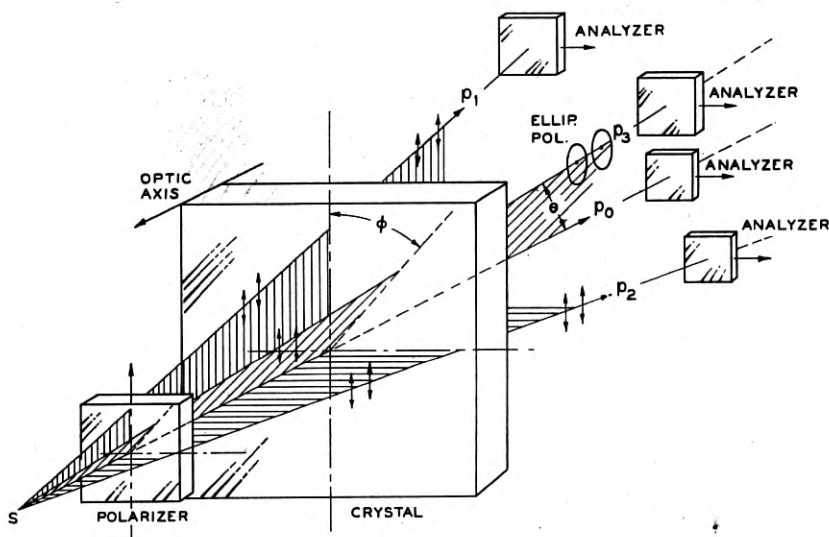
Fig. 2.28—How the polarization changes with propagation direction in a crystal plate

gin with, in order to arrive quickly at some useful conclusions. We will later explain how optical rotation modifies these conclusions.

Consider then the crystal z section shown in Fig. 2.28.   A source s sends monochromatic light through the polarizer which passes only vertical vibrations.   We will assume that the light passes in and out of the crystal without a deviation of path.   Since the vibration is in the plane of z and p_1 (its direction of propagation) the ray does not break up inside the crystal but is propagated as plane polarized light, unchanged.   An analyzer set for vertical extinction could then extinguish this ray.

This is true for propagation from s anywhere in this vertical plane. Also since the vibration is perpendicular to the plane of z and p_2 the ray p_2 does not break up inside the crystal but passes through and out un-

changed. All rays from s in the horizontal plane emerge plane polarized and can be extinguished by an analyzer set for vertical extinction. The ray p_0 is in both these planes so it can be similarly extinguished.

With the ray p_3 the situation is different. Here the vibration is not in the $z\ p_3$ plane so the ray breaks up inside the crystal into two components which travel with different velocities and recombine in or out of phase to give the various degrees of elliptical polarization (including plane and circular). Hence, an eye looking back along p_3, through an analyzer set for vertical extinction, will see light or dark depending on the phase shift N. Now this phase shift for a given thickness of plate is zero along p_0 but increases as θ increases (without changing ϕ; see Fig. 2.28), passing through one integral

$\phi = 0°$

$\phi = 180°$
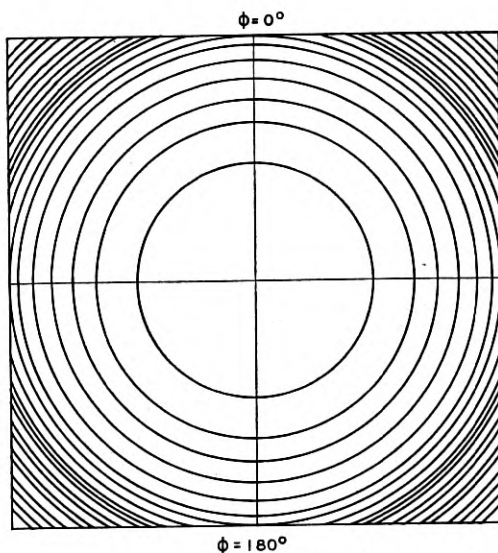
Fig. 2.29—A plot of phase as a function of ϕ and θ

value after another. Therefore, as we allow θ to increase, the eye should see alternate dark and bright regions. Moreover, since the crystal is optically symmetric about z, if ϕ is changed without changing θ, the apparent brightness will not change (except that if $\phi = 0$, 90°, 180° or 270° the field is dark as we previously explained). Consequently, if we could see along all directions at once we would see a pattern of concentric dark rings on a dark cross as shown in Fig. 2.29.

But we *can* see along all these directions at once if we employ a properly placed lens for a lens can converge all these rays to one point where an eye can be placed for viewing.

Thus an eye at e, Fig. 2.30, will see, in the direction $e\ p_0'$, the ray that

started along s p_0. It will see along e p_1' the ray that started along s p_1'. Every point on the lens will have associated with it a different direction in the crystal. Therefore the eye will see a pattern like that of Fig. 2.29. This is the principle of the conoscope. In the conoscope (the name means "conical viewing") the source s is replaced by the image of a source, the image being cast by a lens; see Fig. 2.31. Thus by the use of two similar lenses we get twice as much working space as one lens would give.
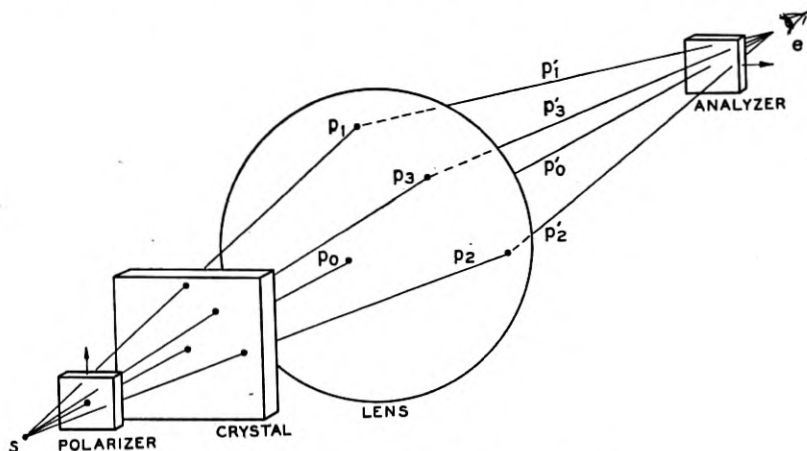
Fig. 2.30—The principle of the conoscope
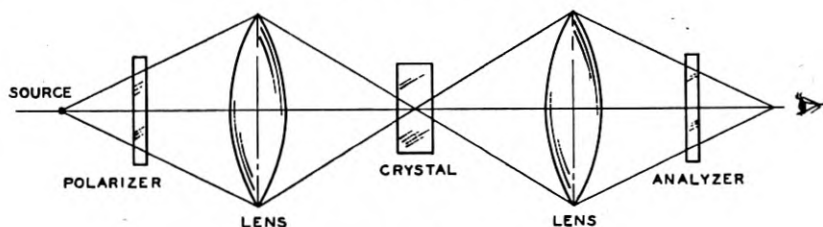
Fig. 2.31—A practical conoscope

Figure 2.32 shows a cross-section of the Western Electric conoscope. The graduated dial shaft goes out through the bottom of the tank to give more working room—older instruments had the shaft overhead and it was in the way. The light source is a mercury arc lamp with filters to isolate the 5461A line. The lenses have a converging power corresponding to f:0.6. The focus is not changed by changes in the refraction of the oil—in fact, the focus is the same with no liquid in the tank as when filled with liquid. This is of some interest for those who might wish to use the instrument for Rochelle salt and accordingly use a fluid of refractive index about 1.495

MERCURY VAPOR LAMP

POLAROID FILTER

REFERENCE GLASS

CARRIAGE

60° VIEWER ATTACHMENT

POLAROID ROTATOR

DIAL

HAND WHEEL
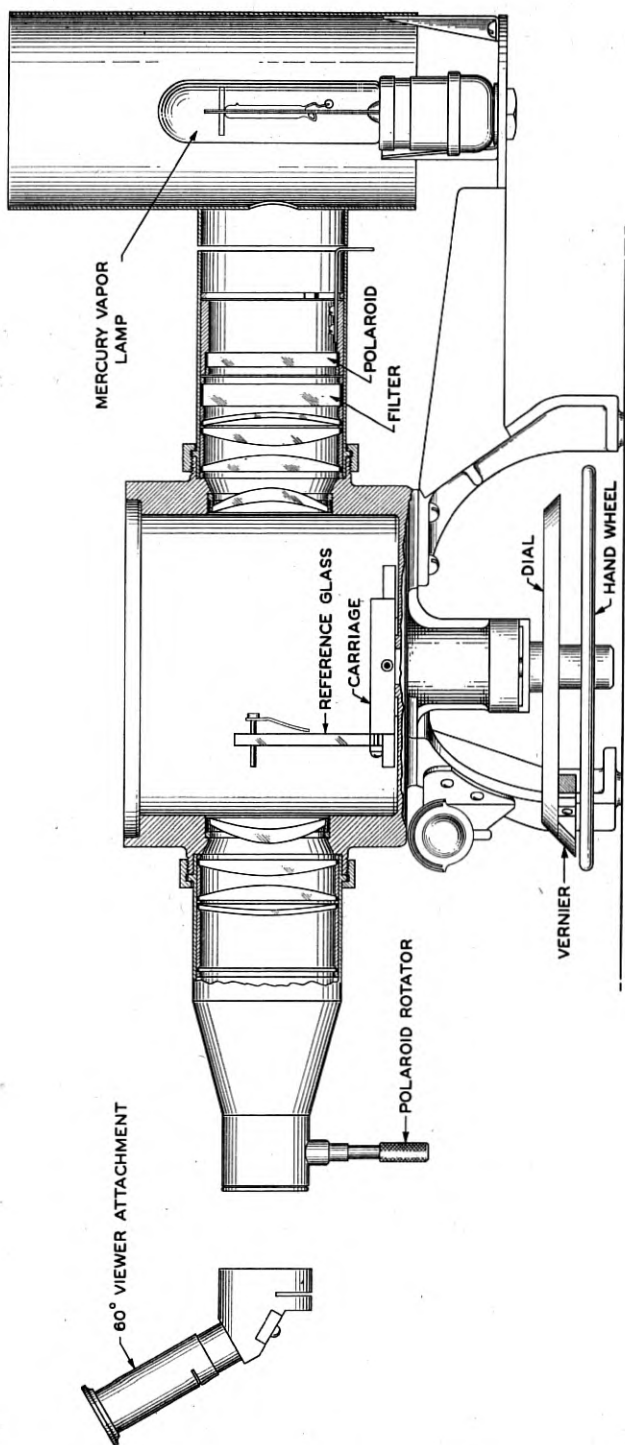
VERNIER

Fig. 2.32—The Western Electric conoscope

instead of the 1.546 of quartz.   The dial is graduated into degrees and a vernier allows readings to tenths of degrees.

If a crystal plate is held against the glass reference surface one may read the angle between the optic axis and the surface normal.   One should occasionally check the instrument (against slippage of the dial) by reversing the crystal and recentering the pattern.   If the readings are not identical, the dial should be adjusted till they are.   Even if the readings are not identical the mean value should be correct.   If one is using the method of ring
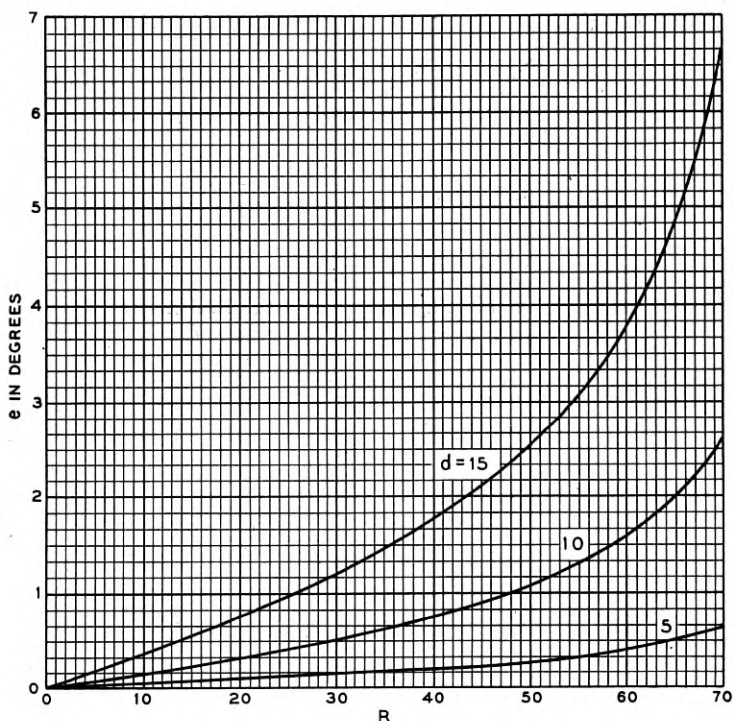
Fig. 2.33—The ring eccentricity correction chart

centering, the correction for eccentricity from Fig. 2.33 should be applied to this mean value.

The carriage may be slid back and forth and for very small crystals the carriage should be placed so that the crystal is near the center of the tank so that very little of the light cone by-passes the crystal.   By the use of a block a thin crystal can be examined by viewing through its edge or length.

The carriage can be removed and "raw" crystals examined.   The optic axis is plainly visible and quite accurate orientations can be made if there is not too much opaque material on the crystal.   Excessive optical twinning

makes a confused pattern but good orientations can be made anyway. A "raw" crystal can be mounted adjustably in a jig that is lowered into the conoscope, the optic axis lined up, the jig transferred to a saw, and sections sawed directly.

Let us turn now to the quantitative analysis of the ring pattern seen in the eye piece when examining a uniaxial crystal. We wish to know the size of the smallest ring in the field, or rather the corresponding angle in the crystal. This first dark ring (analyzer and polarizer crossed) is the result of the slow wave falling *one* wave length behind the fast one. If the plate thickness (Fig. 2.34) is t' the path length in the crystal is

$$t = \frac{t'}{\cos \theta} \tag{2.8}$$
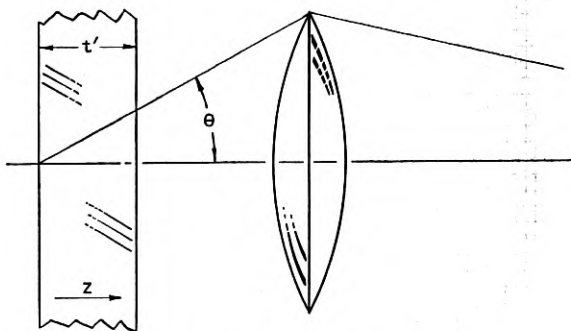
Fig. 2.34—The angle of the smallest ring

This is to be substituted in Eq. 2.7, namely:

$$N = \frac{t}{\lambda} (n_f - n_s) \tag{2.7}$$

Now it can be shown that, quite accurately, at the angle θ from the optic axis:

$$n_s - n_f = .00917 \sin^2 \theta \tag{2.9}$$

where .00917 is the difference in the refractive indices for the ordinary ray and the extraordinary ray for green mercury light traveling at right angles to the optic axis. (These are generally given the symbols n_o and n_e or n_ω and n_ϵ respectively.)

$$N_1 = \frac{t'}{\lambda \cos \theta} \times .00917 \sin^2 \theta = 1$$

and since $\lambda = .000546$ mm. for green mercury light this may be written

$$t' \sin \theta \tan \theta = 0.0595 \text{ mm.} \tag{2.10}$$

whence we solve for the values in this table

convergence	$\theta = 5°$	10°	20°	30°
thickness	$t' = 7.8$	1.94	0.48	0.21

This shows that if we wish to examine thin plates in a conoscope the lenses must be strongly convergent. The conoscope used in the Western Electric has a convergence corresponding to about the 20° entry of the table so it
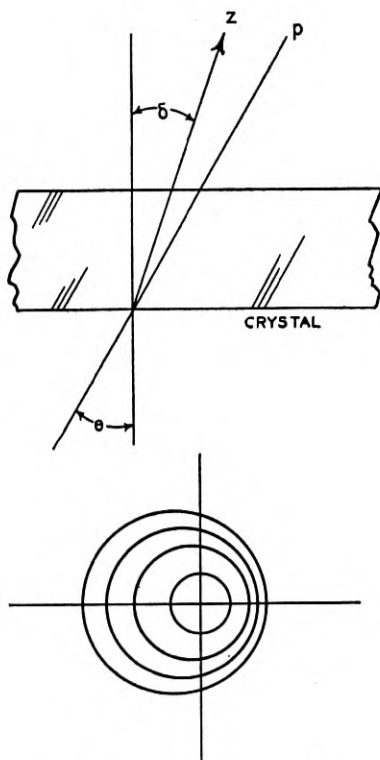

Fig. 2.35—Ring eccentricity

can be used on crystals down to a half millimeter thick—that is for *orientation* studies. In determining *handedness* we remember that this is a question of the rotatory power of quartz causing the rings to expand or contract on rotating the analyzer. Also we said that this rotatory power effectively disappears at 15° from the axis. If no ring is found within 15° of the axis there is no ring capable of expanding or contracting and we cannot test the handedness of such a thin crystal no matter how strong a lens we employ. We can then fall back on the succession of colors shown when we rotate the

analyzer using white light. Red, yellow, blue are observed for clockwise rotation with IRE right-hand quartz.

If the z section is not a true one, Eq. 2.9 will be replaced by one allowing for this error δ:

$$n_s - n_f = .00917 \sin^2 (\theta - \delta)$$

This will make the rings non-concentric and although the "cross" intersection is still the true optic axis the ring centers are not.

Hence if we are tempted to find the optic axis by centering a large sharp ring in preference to a small fuzzy one we find this eccentricity error must be allowed for.

2.8 ROTATION OF THE PLANE OF POLARIZATION

If we have a polarizer and an analyzer set for extinction (Fig. 2.36), then place a thin z section of quartz between them, the field brightens up but can be extinguished again by rotating the analyzer, Fig. 2.37. For the mineralogist's right-hand quartz the analyzer must be rotated 21.7° (yellow light assumed) clockwise to re-establish extinction, counter-clockwise 21.7° for left-hand quartz. The rotation is more for blue light, less for red. If the section is not a perfect z section the rotation is less than this, effectively disappearing at about 15° from the optic axis.

A thick slab can be examined in this way and, due to the color difference in rotation, "rainbows" will be seen in the quartz when held at just the right orientation. These rainbows will follow the contours of the specimen unless both right and left quartz are present in one piece. When this is the case the one kind generally occurs as spike- or blade-like intrusions in the other. It will then cause the rainbows to have sharp, jagged outlines bearing no relation to the specimen contour.

Also, since red, yellow, blue, are here in the order of increasing rotation, if we rotate the analyzer clockwise for right-hand quartz (I.R.E. RHQ) we will pass through best transmission for red, best for yellow and best for blue in that order so that the field will assume these colors in this order.

With uncut stones this examination is best made under an immersion fluid. The *inspectoscope* is made for this work. We spoke of the rotation of the plane of polarization and its complicating of the issue for the conoscope. Due to this the field at the center is not dark when the analyzer and polarizer are crossed. Also if we rotate the analyzer clockwise the rings of the pattern either expand or contract according to whether the crystal is right-hand quartz or left-hand quartz (IRE definition).

A different kind of pattern is visible in the conoscope when viewed perpendicular to z, a double set of hyperbolae as shown in Fig. 2.38. This pat-

tern has been used to check the orientation but much grief has ensued due to not recognizing one of its properties. This property is that, if *z* does not lie parallel to the crystal boundary the center of the pattern is not perpendicular to the optic axis and a rather involved correction must be used. This correction reduces the actual angle to about half the observed value.

This conoscope is an immersion instrument. The fluid is chosen to have an index of refraction to match the "ordinary" one for quartz. When this is done light is not bent in passing between fluid and quartz. When the fluid does *not* match there is a bending and all readings are subject to a
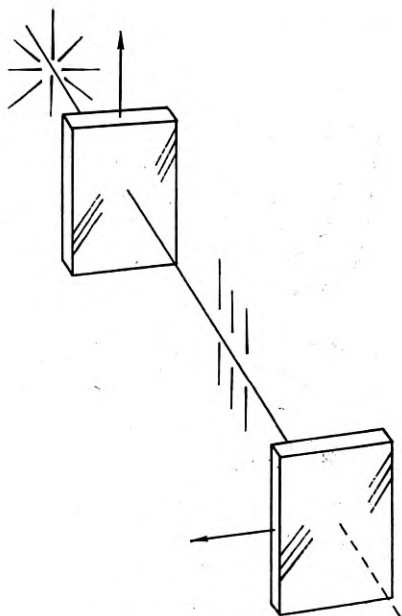
Fig. 2.36—Crossed polarizers   Fig. 2.37—Rotation of the plane of polarizer
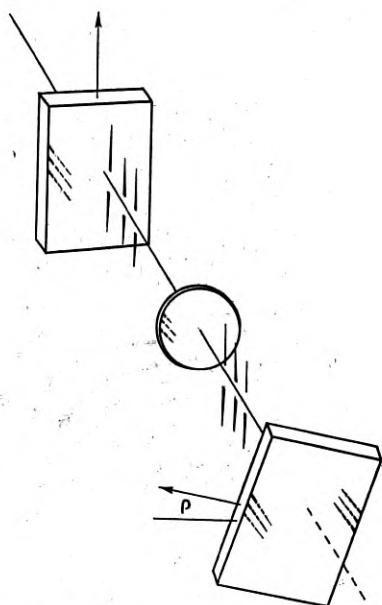
correction. For example if we measure the angle of an AT plate in a fluid that is too low by .0048 (since n_0 for quartz in green mercury light is 1.5462 this fluid has $n = 1.5414$), we will get a reading that is too high by a quarter degree (the 35° angle will then appear low). A temperature rise of 12° C will lower the relative refractive indices by this amount.

Also the more perfectly the fluid matches, the more nearly will the rough quartz surface disappear and seem smooth and clear. This greatly enhances the sharpness of the rings.

The *refractoscope* (Fig. 2.39) was designed by G. W. Willard to tell when the

match is good.   It uses the elimination of the bending as a test for refractive
match.   It also demonstrates the existence of two velocities in quartz, for
two images are seen where a glass prism would cause but one.   Also by
viewing through an analyzer we see that the two images are caused by plane
polarized light, the polarization planes being mutually perpendicular.

If the fluid has an index lower than that of the prism the rays will bend
towards the base of the prism.   For this reason, light that reaches the eye e
from s must travel by the path $s\ q_0 p_0 e$ for the ordinary ray, $s\ q_\epsilon p_\epsilon e$ for the
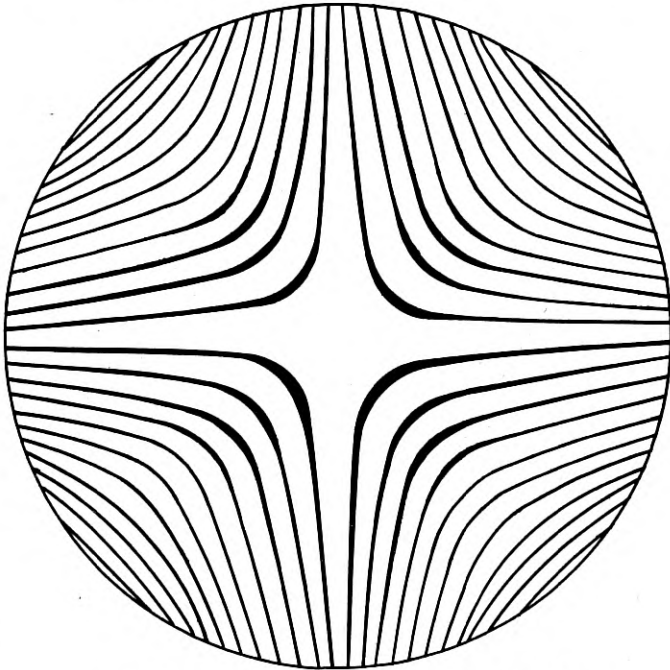
Fig. 2.38—Conoscope view normal to the optic axis

extraordinary ray.   Hence the slit as seen through the prism will appear at
s_0' for the ordinary ray, at s_ϵ' for the extraordinary while the slit as seen
directly alongside the prism will appear at s.   If the fluid index matches
the quartz prism for the ordinary ray this ray will be unbent at p_0 and q_0 so
that s_0' will appear as a continuation of s.

If the fluid index is too high the image s_0' will appear to the left of s with
s_ϵ' still to the right of s_0'.

As the refractive indices of quartz for the ordinary and the extraordinary
ray differ by .009 the apparent separation of s_0 and s_ϵ' represents .009 and

can be used to judge the difference between the liquid index and the quartz ordinary index.

The liquid can be adjusted by placing a cap face of a crystal on the reference glass and setting the dial to read 51.8° the liquid being then blended to center the pattern.

If the refractive index of the fluid is low by an amount L the observed reading R must be corrected by adding to it an amount ϵ.

Since

$$(N_q - L) \sin R = N \sin (R - \epsilon)$$

we can compute ϵ. The accompanying refraction correction nomograph was computed from the above equation. If we know that the fluid in the conoscope is high, by an amount H, and we wish to know the correction to be applied to a conoscope reading R we locate H on the diagonal line HL and locate R on the horizontal line R. We join these two points with a straight line and read the scale $\epsilon - \epsilon$ where this straight line crosses the curved line $\epsilon - \epsilon$. This value tells the size of the correction and whether to add it to R or subtract it from R. Conversely, if we wish to find how closely the index of the fluid must be held in order to have the correction less than say $\frac{1}{4}°$ at a reading of say $R = 50°$ we join the points $R = 50°$ and $\epsilon = \pm \frac{1}{4}°$ and find $H = \pm 0.005$. A ten-inch-long lucite strip with a straight line ruled on it is a convenient tool with which to read this nomograph.

We now inquire as to whether the refraction correction can be made to annul the ring eccentricity correction. In the appendix it is shown that this is done if $H = -.530 \tan^2 d$ where $2d$ is the distance between the vertical reticule lines.

Experimentally it is easy to achieve this balance by using a cap face parallel slice. With the cap face against the reference glass and the dial reading 51.8° the fluid is blended to make a single ring tangent to both reticule lines. When this is done for $d = 10°$ the fluid should have a refractive index of 1.5228 and the residual errors should be less than 2 minutes for R not over 60°.

2.9 IMMERSION FLUIDS

In order to match the refractive index of quartz we blend a substance which has an index that is too large with one that has an index that is too small. Such blended substances should be liquid at room temperature and hence should be perfectly mutually soluble. They should have low vapor pressure so that they do not evaporate quickly and should be harmless to the operator. Also they should be nearly colorless and clear. They should be fluent enough to be easily drained from the crystal and should have a

sufficiently high flash point that they would not present a fire hazard. The odor should not cause distress and finally the cost must be reasonable.

Dr. G. T. Kohman of the Bell Telephone Laboratories has prepared a list of such substances that can be mixed, any high one with any low one to
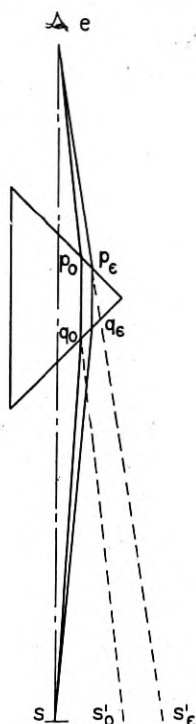

Fig. 2.39—The refractoscope

obtain a fluid satisfactory in all these respects. The following mixtures of substances are taken from his data.

Substance	Refractive Index	Mixture Parts by wt.	Density	Flash Point	Supplier
Dimethyl phthalate	1.51	73.9	1.193	255°F	Monsanto Chem. Co.
α monochlor naphthalene	1.63	26.1	1.194		Bakelite Co.
Dimethyl phthalate	1.51	73.9	1.193	285°F	Monsanto Chem. Co.
Dichlor naphthalene (solid at toom temp.)	1.63+	26.1	1.30		Hooker Chem. Co.
Decalin	1.467	35.3	0.895	170°F	Dupont Co.
Dowtherm	1.586	64.7	1.1		Dow Chem. Co.
Kerosene				170°F	

An immersion fluid for Rochelle salt can be made by mixing decalin with any of the other substances. For example *b* mixture of 34 parts of dimethyl phthalate and 66 parts of decalin should give the necessary index 1.495.

2.10 APPENDIX

THE RING ECCENTRICITY CORRECTION

Referring to Fig. 2.41, we see that, at an angle α, from the optic axis towards the plate thickness direction the phase relation is, by Equations (2.7) and (2.9):

$$N_1 = \frac{.00917 \; t' \sin^2 \alpha_1}{\lambda \cos(\delta - \alpha_1)}$$
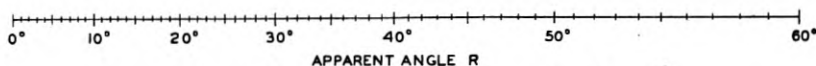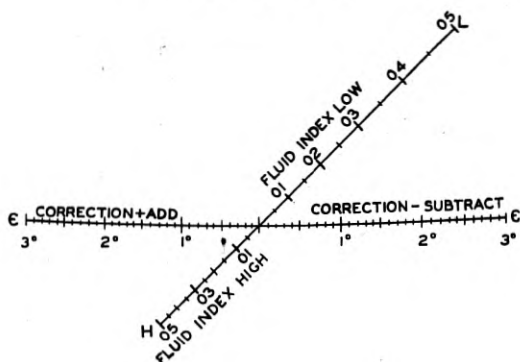
Fig. 2.40—Refraction correction nomograph

while at an angle α_2 away, it is:

$$N_2 = \frac{.00917 \; t' \sin^2 \alpha_2}{\lambda \cos(\delta + \alpha_2)}$$

Whence, if these are equal, we have:

$$\frac{\sin^2 \alpha_1}{\cos(\delta - \alpha_1)} = \frac{\sin^2 \alpha_2}{\cos(\delta + \alpha_2)} \qquad (2.11)$$

These points, in the conoscope field, being of equal phase are parts of the same ring, and if matched to a pair of reticule lines, the optic axis is off their center line by an angle e, where

$$e = \frac{\alpha_1 - \alpha_2}{2}$$

If the separation of the reticule lines corresponds to an angle $2d$, we see that
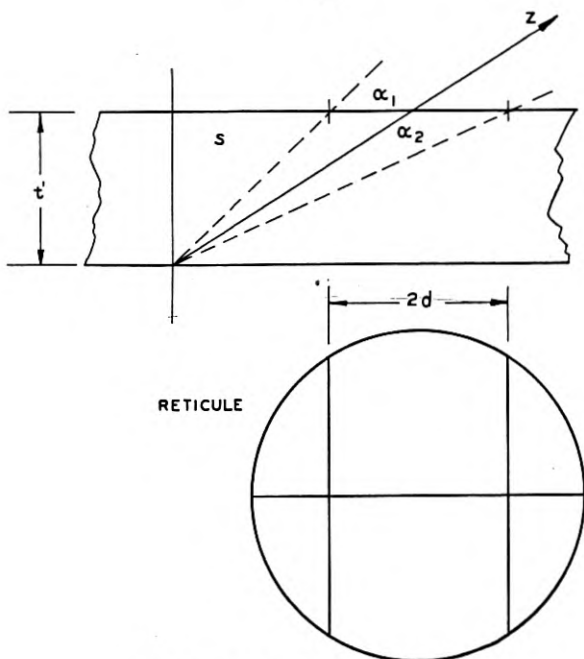
$$d = \frac{\alpha_1 + \alpha_2}{2}$$

Fig. 2.41—Off center cross hairs

so that

$$\alpha_1 = d + e \quad \text{and} \quad \alpha_2 = d - e$$

The reading of the dial will be, at this match

$$R = \tfrac{1}{2}(\delta + \alpha_2 + \delta - \alpha_2) = \delta - \frac{\alpha_1 - \alpha_2}{2}$$

Hence we have

$$\frac{\sin(d + e)}{\sin(d - e)} = \sqrt{\frac{\cos(R - d)}{\cos(R + d)}} \tag{2.12}$$

For a given value of d, we can plot the values of e as a function of δ as given by (2.12).   This plot is a chart of corrections to be added to the readings R to find the true angle δ.

Examination of Equation (2.12) shows us that the correction e is independent of the thickness t' and even of the birefringence; hence, the chart could serve for all uniaxial crystals.   Equation (2.12) can be given an approximate solution:

$$e = 1820 \tan^2 d \tan R \text{ minutes}$$

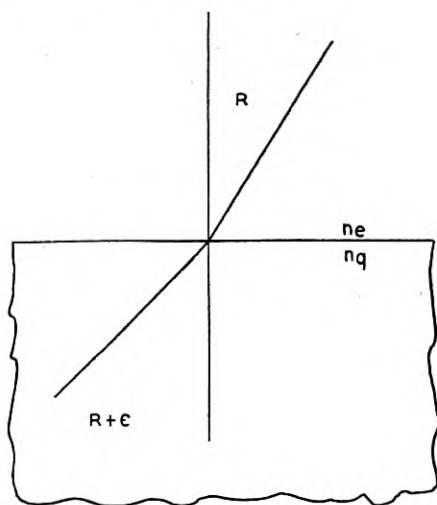$$e = 30.3 \tan^2 d \tan R \text{ degrees} \tag{2.13}$$

Fig. 2.42—Refraction at a surface

For R not more than 60 degrees and d not more than 15 degrees the error in e is not more than 5 minutes.   Figure 2.33 is a chart of these corrections computed from the more exact equation (2.12).

ANNULLING THE RING ECCENTRICITY CORRECTION BY MEANS OF THE
REFRACTION CORRECTION

The difference in quartz index and liquid index is

$$H = N_q - N_\ell$$

and by the law of refraction:

$$N_\ell \sin R = N_q \sin (R + \epsilon)$$

whence

$$-H = N_q \frac{\sin \epsilon}{\tan R} - (1 - \cos \epsilon)$$

and if ϵ is small

$$H = -\frac{N_q \sin \epsilon}{\tan R}$$

From Eq. 2.13, $e = 0.530 \tan^2 d \tan R$ radians, and putting this in the equation for H we find that the correction for ring eccentricity approximately annuls the correction for refraction if

$$H = -.530 \tan^2 d$$

For $d = 10°$ this gives $H = -.0255$, that is, a fluid index of 1.5207.

A Note on the Transmission Line Equation in Terms of Impedance

By J. R. PIERCE

INCREASED familiarity derived in handling Maxwell's equations, especially in connection with problems arising at very high frequencies, has resulted in a variety of forms for expressing certain laws and behavior. Especially, work by Schelkunoff in extending the impedance concept[1] shows that impedance can be quite as general and exact a means for expressing electromagnetic relations as are current, voltage, electric and magnetic fields, and vector and scalar potentials.

In reformulating certain problems in terms of impedance the content and ultimate solution must of course be equivalent. There may, however, be a considerable change of procedure and sometimes a simplification. For instance, in many cases a single impedance condition can replace the usual two boundary conditions for voltage and current.

One very simple case in which it is perhaps easiest to deal directly with impedance is in the derivation of the transmission line equation on a distributed- constant- basis. In the usual derivation, two linear second order differential equations are obtained, one for voltage and one for current. The impedance, in terms of which the engineer expresses many of his results, is obtained as a ratio from solutions for voltage and current. In treating the transmission line from the impedance point of view, without dealing with currents and voltages, a first order non-linear differential equation in terms of impedance and distance is obtained. This impedance equation is a Ricatti equation and could be obtained from the usual line equations. It is simpler, however, to derive it directly.

As the principal interest of such a treatment lies in the method and in the fact that the line may be tapered, rather than in losses, the derivations will be carried out for lossless lines. Losses can be taken into account by allowing the inductance per unit length, L, and the capacitance per unit length, C, to become complex quantities.

Consider the section of line dx long, shown in the figure, having an inductance $L\, dx$ and a capacitance $C\, dx$. We can write immediately

$$Z_x + dZ = Z_{x+dx}$$
$$= j\omega L\, dx + \frac{1}{j\omega C\, dx + 1/Z_x}$$
$$= Z_x + j\omega[L - CZ_x^2]\, dx. \tag{1}$$

---

[1] "The Impedance Concept and Its Application to Problems of Reflection, Refraction, Shielding, and Power Absorption," *B.S.T.J.* Vol. 17, pp. 17–48, January, 1938.

Dropping the subscript x, the differential equation for the line in terms of the impedance Z may be written[2]

$$R \frac{dZ}{dx} = j \frac{\omega}{v} (R^2 - Z^2) \tag{2}$$

$$R = (L/C)^{1/2} \tag{3}$$

$$v = (LC)^{-1/2} \tag{4}$$

R is the nominal characteristic impedance, and v is the nominal phase velocity, which is constant for many tapered lines with the same dielectric material separating the conductors throughout their length. In such lines, if the dielectric is air or vacuum, v is c, the velocity of light.

It should not be surprising that (2) is of the first order. Although there are two boundary conditions, the impedances terminating the right and left ends of the line, there are two impedances, that looking toward the right and that looking toward the left. The impedance looking toward the right
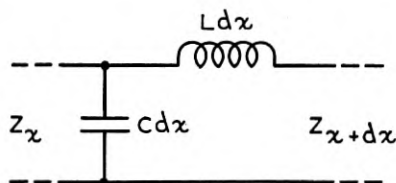

Fig. 1

is unaffected by the left end termination, and that looking toward the left is unaffected by the right end termination.

As R is real, it may be seen from (2) that the only case in which the impedance can equal the nominal characteristic impedance R at all points is for R constant. This tells us that the characteristic impedance of any lossless tapered line is complex. For very gradually tapering lines the characteristic impedance differs from the nominal characteristic impedance principally by a small imaginary component.

The simplest solution of (2) is of course that for a uniform line, with R a constant which will be called R_o. In this case (2) can be integrated directly, giving the familiar result

$$\frac{Z}{R_0} = \tanh (j\omega x/v + K) \tag{5}$$

[2] It is interesting to note that the equation for admittance Y can be obtained by replacing Z by Y and R by $(1/R) = G$ in (2).

Dr. L. A. MacColl has pointed out to the writer that (2) is the same as the electrostatic electron optical equation for paraxial rays. To reduce (2) to the standard form:

$$-\frac{j\omega\, dx}{Rv} = dz \tag{6}$$

$$-R^2 = H(z) \tag{7}$$

$$\frac{dZ}{dz} = H(z) + Z^2 \tag{8}$$

The electron optical equation for paraxial rays is

$$\frac{d\Gamma}{dz} = \frac{3}{16}\left[\frac{V'(z)}{V(z)}\right]^2 + \Gamma^2 \tag{9}$$

$$\Gamma = C - \frac{V'(z)}{4V(z)} \tag{10}$$

Here z is distance along the axis, $V(z)$ is potential on the axis, and C is convergence, or the inverse of the focal distance.

It would seem, then, that from each solution of an electron optical problem, a solution of a tapered line problem could be found, and vice versa.

While it cannot be claimed that anything new has entered the transmission line equation in expressing it in terms of impedance, it does seem that the approach may be stimulating in uncovering hitherto neglected material and analogies.

Abstracts of Technical Articles by Bell System Authors

Electronics in Telephony.[1] FRANK A. COWAN. The historical development of the use of electronic devices by the telephone system is reviewed, showing how long distance telephony has grown with the increased use of, and improvements in, electronics. The number of telephone repeaters has grown from 16 in 1908 to 123,000 in 1942 and carrier circuit mileage has grown from 2,000 in 1920 to 2,300,000 in 1942, while copper usage per circuit mile has decreased from 400 pounds in 1910 to less than 70 in 1942.

A transcontinental telephone connection has grown from an open-wire circuit with a total loss, less repeaters, of less than 75 db (1915) to a present day cable circuit operating at carrier frequencies, which may have a total loss of over 10,000 db. The problem of matching enormous amplifications to compensate for huge losses with a precision of one or two db was a difficult one, which was solved by electronic techniques. The amplification necessary to compensate for the high losses on the cable layout may entail some 200 repeaters utilizing a total of more than 600 vacuum tubes in tandem. The automatic regulation and control of the amplification is accomplished by electronic devices, giving to the present day circuits a stability unequalled in the days before the vacuum tube.

There is available, except for the War, radiotelephone service to 83 foreign countries and overseas areas, and ocean liners at sea, and to boats in coastal and inland waters.

Such widespread dependence on vacuum tubes has stimulated research and design to achieve long life and a high degree of uniformity, stability and reliability. Among interesting future possibilities are transoceanic cables, the use of higher frequencies providing broader bands and larger numbers of circuits over a given path, and further application of remote and unattended stations.

Deionization Considerations in a Harmonic Generator Employing a Gas-Tube Switch.[2] WILLIAM G. SHEPHERD. A description is given of an experimental investigation of the properties of a thyratron operating as a high-frequency switch in a circuit which permitted the generation of a wide band of harmonics. The experiments indicate that there is an operating frequency below which no difficulties in deionization occur and above

[1] *Electronics*, March 1943.
[2] *Proc. I.R.E.*, February 1943.

266

which stable operation requires that the grid potential fulfill certain conditions dependent upon the frequency, wave form of the grid voltage, and circuit constants. It has been found possible to operate certain standard thyratrons at switching frequencies as high as several hundred kilocycles per second. For these higher frequencies the deionization of the tubes is incomplete but normal switching behavior is obtained.

Contributors to this Issue

WALTER L. BOND, B.S. in Physics, Washington State College, 1927; M.S. 1928. Member of Technical Staff, Bell Telephone Laboratories 1928–. Studied at Columbia University, New York University, and Stevens Institute. Engaged primarily in studies of the physical properties of crystals.

L. A. MacCOLL, A.B., University of Colorado, 1919; A.M., Columbia University, 1925; Ph.D., Columbia University, 1934. Engineering Department, Western Electric Company, 1919–25; Bell Telephone Laboratories, 1925–. Dr. MacColl is engaged in research in mathematics and mathematical physics.

W. P. MASON, B.S. in Electrical Engineering, University of Kansas, 1921; M.A., Columbia University, 1924; Ph.D., 1928. Bell Telephone Laboratories, 1921–. Dr. Mason has been engaged in investigations on carrier systems and in work on wave transmission networks both electrical and mechanical. He is now head of the department investigating piezoelectric crystals.

J.R. PIERCE, B.S. in Electrical Engineering, California Institute of Technology, 1933; Ph.D., 1936. Bell Telephone Laboratories, 1936–. Engaged in study of vacuum tubes.