PHYSICAL REVIEW B VOLUME 1, NUMBER 6 15 MARCH 1970

Symmetry Considerations and the Vanadium Dioxide Phase Transition

J. R. Brews
Bell Telephone Laboratories, Murray Hill, New Jersey 07974

(Received 2 October 1969)

Some consequences of the symmetry change associated with the vanadium dioxide metal-semiconductor transition are derived. A free-energy expansion in the ionic displacements has been used. The effect of a ⟨101⟩ zone-edge phonon instability and strain effects have been included. The symmetry arguments have been carried out using multiplier representations of the symmetric phase space group D_{4h}^{14}. It is found that a first-order transition from the symmetric phase to the monoclinic phase, C_{2h}^{5}, is compatible with such a free-energy expansion. Domain structure is predicted, which agrees with observations available. The dependence upon the spontaneous distortion of the dielectric tensor is derived. While the results are compatible with only one ⟨101⟩ zone-edge phonon becoming unstable, the possibility of more than one such phonon participating in the transition cannot be ruled out.

I. INTRODUCTION

A NUMBER of microscopic models for the nonmagnetic vanadium dioxide metal-semiconductor first-order phase transition have been proposed.[1–3] These models have focused attention upon the striking conductivity change at the transition temperature and have treated the accompanying symmetry change[4,5] from space group D_{4h}^{14} to C_{2h}^{5} as a secondary effect. It is the purpose of this paper to consider the possible consequences of this symmetry change, particularly as it affects domain structure and coupling of the transition mechanism to an external electric field.

The previously proposed microscopic models of the transition[1–3] are inadequate for this purpose. First, these models are *microscopic* and so aim at an understanding more ambitious than required for the symmetry considerations we have in mind. Quantitative versions of these models which could predict crystal symmetry would be extremely complicated mathematically. Secondly, there is growing evidence[6] that a basic assumption of these models, namely, that of a very narrow conduction band width, does not apply to vanadium dioxide. Some other difficulties with these models are discussed in Ref. 6.

An alternative to the microscopic model approach is the free-energy expansion model so successfully employed for ferroelectrics.[7] Such an approach emphasizes the lattice dynamical aspect of the transition,[8] that is, the change in space group at the transition. Being more phenomenological than the microscopic models, this approach requires less detailed assumptions about the microscopic mechanisms driving the transition. Of course, before the transition can be considered as understood a microscopic derivation of the free energy expansion is mandatory. However, at an intermediary level of understanding such as now exists, the phenomenological expansion can serve as a convenient starting point for a symmetry analysis.

The question of validity of a free-energy expansion for a first-order transition is somewhat unclear. The microscopic models have seldom been carried through to a free-energy expression. An exception is the one-dimensional Adler-Brooks crystal distortion model[7] which predicts a free-energy nonanalytic in the crystal distortion. This nonanalyticity disappears[1] for the wider bands that accumulating evidence suggests for vanadium dioxide.[6] Other attempts to justify free-energy expansions for lattice instabilities are based upon a modified perturbation theory treatment of anharmonic effects.[8,9] These treatments assume that perturbation theory will work and, to be practical, they must also assume that only a few terms are needed (i.e., small displacements of the ions from their symmetric phase equilibrium positions). More general theoretical arguments for second-order phase transitions suggest that validity of a free-energy expansion is not dependent on the detailed nature of the interaction causing the transition but only upon the range of this interaction.[10,11] Furthermore, systems exhibiting first-order phase transitions well described by free energy expansions do exist.[7] Therefore, we take the view that it is reasonable to adopt a free-energy expansion, since theoretical results suggest its validity for a wide variety of systems, and since physical systems do exist where such expansions work.

If one were to attempt to construct a microscopic lattice dynamical model of the transition, the number of rutile phase phonons participating in the transition would be of interest. The bearing of symmetry arguments upon this point is also considered in this paper. However, for the phenomenological determination of the free-energy expansion the number of unstable

---

[1] D. Adler, in *Solid State Physics*, edited by F. Seitz *et al.* (Academic Press Inc., New York, 1968), Vol. 21, p. 1.

[2] N. F. Mott, Phil. Mag. **20**, 1 (1969).

[3] H. Frohlich, in *Quantum Theory of Atoms, Molecules, and the Solid State*, edited P. Lowdin (Academic Press Inc., New York, 1966), p. 465.

[4] S. Westman, Acta Chem. Scand. **15**, 217 (1961).

[5] G. Andersson, Acta Chem. Scand. **10**, 623 (1956).

[6] C. N. Berglund and H. J. Guggenheim, Phys. Rev. **185**, 1022 (1969).

[7] A. F. Devonshire, Phil. Mag. **40**, 1040 (1949).

[8] P. C. Kwok and P. B. Miller, Phys. Rev. **151**, 387 (1966).

[9] A. A. Maradudin, in *Symposium on Ferroelectricity*, edited by E. F. Weller (Elsevier Publishing Co., New York, 1967), p. 72.

[10] L. P. Kadanoff *et al.*, Rev. Mod. Phys. **39**, 395 (1967).

[11] D. J. Thouless, Phys. Rev. **181**, 954 (1969).

phonons is not of interest, but rather the number of different symmetries of unstable phonons (see Sec. VI).

II. MATHEMATICAL STATEMENT OF PROBLEM

The background to free-energy expansions has been previously discussed by Landau and Lifshitz,[12] by Lyubarski,[13] and by Birman.[14] However, an approach which is very suited to a discussion of domain types and which provides a clearer understanding of what have been called "virtually possible" changes in crystal symmetry[15] (see Sec. IV) is based upon the variety of degenerate ground states available to the crystal in its distorted phase. Domain structure in a macroscopic sample results when different regions of the crystal adopt different choices for ground state.

Such a multiplicity of ground states can be shown to exist if, as is the case for vanadium dioxide (see Appendix), the distorted phase differs from the symmetric phase only by a loss of some symmetry operations.[16] The full symmetry of the symmetric phase reflects itself in these degenerate ground states. Every symmetry operation which is lost in a particular distorted state generates from this particular distorted state another one. What is more, the symmetry operations preserved in any particular distorted state must form a subgroup of the symmetric phase symmetry group. The free energy itself remains invariant in both phases to symmetry operations of the symmetric phase-space group as follows from the definition

$$F=-kT\ln \mathrm{Tr}e^{-H/kT}, \qquad (1)$$

where H is the crystal Hamiltonian which includes the interactions which ultimately cause the transition to the distorted phase, but not any minor perturbations which would cause one distorted ground state to be preferred over another.

Suppose that from these various ground states a linearly independent set is chosen $\{\psi_\beta\}$ and that all the ground states $\{R\psi\}$ are expressed in terms of the $\{\psi_\beta\}$ (R= a symmetry operator of H). Where a crystal distortion is involved and no magnetic ordering occurs each ψ_β can be characterized by its ionic configuration, e.g., by specifying the ionic displacements, $u_\beta(n;ki)$, from the symmetric phase equilibrium positions. Here, $u_\beta(n;ki)$ is the displacement in the nth cell of the kth ion along the ith coordinate axis for βth distorted ground state. Because the $\{u_\beta\}$ transform into linear combinations of themselves under the symmetry operations of the symmetric phase-space group, they form a basis for a representation of this space group. If this representation is reduced to its irreducible components, then each u_β becomes a linear combination of the basis vectors $\{f_{qj}\}$, say, of the irreducible representations $\{q\}$,

$$u_\beta=\sum c_{\beta;qj}f_{qj}. \qquad (2)$$

If the distorted state symmetry is known, Eq. (2) restricts the irreducible representations q which may occur in the expansion of u_β since u_β is invariant under those symmetry operations R which are also symmetry operations of the distorted-phase space group.

The free energy F is an invariant of the symmetric phase space group. To make use of the invariance condition, F is expanded in powers of the $\{c_{\beta;qj}\}$:

$$F=F_0+\sum F_1(\beta;qj)c_{\beta;qj} +\sum\sum F_2(\beta\beta';qj,q'j')c_{\beta;qj}c_{\beta';q'j'}\text{, etc.} \qquad (3)$$

The mathematical problem then divides into two parts. The first part is the determination of which irreducible representations q are allowed in the expansion of Eq. (2). The second part is the determination of the linear, quadratic, etc., invariants involving $\{c_{\beta;qj}\}$, $\{c_{\beta;qj}c_{\beta';q'j'}\}$, etc., which can be formed from these allowed representations and the reexpression of Eq. (3) in terms of these invariants.

Even granting the validity of such an expansion, Eq. (3) is equivalent to the free energy of Eq. (1) only with the understanding that the $\{c_{\beta;qj}\}$ are not arbitrary parameters but are fixed by the requirement that at any given temperature they take on the values which minimize the F of Eq. (3). These minimization conditions result in equations giving the T dependence of the $\{c_{\beta;qj}\}$. For example, in the Kwok-Miller analysis of displacive ferroelectrics[8] the $\{c_{\beta;qj}\}$ are the thermal expectation values of the normal-mode amplitudes.

III. TRANSLATION GROUP RESTRICTIONS ON ALLOWED REPRESENTATIONS

The simplest restrictions on the irreducible representations entering Eq. (2) stem from the requirement that u_β be invariant under translations of the distorted phase. Such translation restrictions have been discussed in general by Haas[17] and by Miller and Kwok.[18] These authors restrict themselves to the case when the distorted phase unit cell is an integer multiple of the symmetric phase unit cell. This restriction is the same as the assumption that the translation group of the distorted phase is a subgroup of the translation group of the symmetric phase, an assumption also made in Sec. II of the present paper. It is then found that the only irreducible representations that are allowed in Eq. (2) are those characterized by wave vectors which are in the symmetric phase Brillouin zone and also are distorted phase reciprocal-lattice vectors. Here these restrictions

---

[12] L. D. Landau and E. M. Lifshitz, *Statistical Physics* (Addison-Wesley Publishing Co., Inc., Reading, Mass., 1958), p. 430.

[13] G. Ya. Lyubarski, *The Applications of Group Theory in Physics* (Pergamon Press, Ltd., Oxford, 1961), p. 121, *et seq.*

[14] J. L. Birman, in *Symposium on Ferroelectricity*, edited by E. F. Weller (Elsevier Publishing Co., New York, 1967), p. 20.

[15] Reference 12, p. 452.

[16] R. D. Mattuck and B. Johansson, Advan. Phys. **17**, 509 (1968).

[17] C. Haas, Phys. Rev. **140**, A863 (1965).

[18] P. B. Miller and P. C. Kwok, Solid State Commun. **5**, 57 (1967).

will be related to the various degenerate ground states labeled by β.

It follows from the assumption that the distorted-state symmetry operators form a subgroup of the symmetric phase-space group that each u_β is associated with some particular way of choosing the subset of monoclinic-space-lattice vectors from the set of rutile-space-lattice vectors. For example, it is shown in the Appendix that one choice for the general translation of the monoclinic phase $\mathbf{x}(m)$ is

$$\mathbf{x}(m)=m_3\mathbf{a}_1(r)+m_2\mathbf{a}_2(r)+(2m_1-m_3)\mathbf{a}_3(r)\,, \qquad (4)$$

where the $\{m_i\}$ are integers and the $\{\mathbf{a}_i(r)\}$ are basis vectors for the rutile lattice. Any other ways of imbedding the monoclinic structure in the rutile structure must be obtained from Eq. (4) by some rutile symmetry operation, since all the distorted degenerate ground states are interrelated by such operations. Hence the reciprocal lattices for the various ground states are also related by symmetry operations. This leads to the conclusion that if one of the wave vectors for a given star of wave vectors appears in the expansions of Eqs. (2) and (3), so must they all.

The Kwok-Miller theorem on allowed wave vectors when applied to the distorted state of Eq. (4) produces

$$\mathbf{q}=0 \qquad (5)$$

and

$$\mathbf{q}=\pm\tfrac{1}{2}(\mathbf{b}_1\pm\mathbf{b}_3)\,, \qquad (6)$$

where $\{\mathbf{b}_i\}$ are the rutile reciprocal lattice vectors. The wave vectors appearing in Eqs. (5) and (6) represent only two stars. The $\mathbf{q}=0$ wave vector forms its own star since, of course, it is not related to the $\langle 101\rangle$ wave vectors by any symmetry operation. The wave vector $\frac{1}{2}(\mathbf{b}_1+\mathbf{b}_3)$ is related to all the other wave vectors of Eq. (6) by the symmetry operations $(\sigma_h/001)$, $(\sigma_{vx}/11\bar{1})$, and $(C_{2y}/110)$ (see Table I for notation). Hence all the wave vectors of Eq. (6) belong to the same star. However, the symmetry operation $(C_4/001)$, a fourfold screw axis parallel to the rutile c axis, carries the wave vectors of Eq. (6) into

$$\mathbf{q}=\pm\tfrac{1}{2}(\mathbf{b}_2\pm\mathbf{b}_3)\,. \qquad (7)$$

Hence these also must be allowed for some other members of the set $\{u_\beta\}$ of monoclinic distorted ground states. Another unit cell, related to the first by $(C_4/001)$, also describes a crystal structure with space group $C_{2h}{}^5$. If this unit cell is used the roles of $\mathbf{a}_1(r)$ and $\mathbf{a}_2(r)$ are interchanged in Eq. (4) and Eq. (7) results directly from Eq. (4) thus modified. The wave vectors of Eq. (6) and Eq. (7) exhaust the $q=\langle 101\rangle$ star.

From the above discussion it should be clear that none of the other distorted ground states can lead to any other star than the two just found. Only the $\mathbf{q}=0$ and the $\mathbf{q}=\langle 101\rangle$ stars occur in the expansions of Eq. (2) and Eq. (3).

IV. RESTRICTIONS IMPOSED ON REPRESENTATIONS BY OTHER SYMMETRY ELEMENTS

There are a number of space groups compatible with any given translation group, so that specifying the stars of the wave vectors of the irreducible representations allowed in Eq. (2) is not sufficient to insure that the corresponding free-energy expansion will describe only the monoclinic phases with space group $C_{2h}{}^5$. One can treat *all* the allowed space groups compatible with the translations of Eq. (4) (and the symmetry related translations) and also compatible with the particular ionic configuration found in rutile by considering *all* ionic displacements compatible with the translations of Eq. (4) (and the symmetry related translations) and the ionic configuration of rutile. This is the approach taken by Lyubarski.[13] The resulting occurrence of space groups other than the actually realized space group ($C_{2h}{}^5$ in the present case) has been referred to as the occurrence of "virtually possible" symmetry changes.[15] The particular representations allowed in Eq. (2) by the space group $C_{2h}{}^5$ then can be obtained easily by further restriction to the subset of the general displacements which is invariant under $C_{2h}{}^5$, viz., the $\{u_\beta(n;ki)\}$.

To carry out this approach a method similar to that used in analyzing molecular vibrations, but generalized to allow wave vector dependence, will be adopted. First, a complete set of ionic displacements compatible with a given wave vector is defined, and the transformation matrices of this set are constructed. These form a reducible representation of the group of the given wave vector. These matrices are then related by a phase factor to another set of matrices which form a reducible, unitary, *multiplier* representation of the *point* group of the given wave vector. This multiplier representation is then reduced and the corresponding basis vectors for each contributing irreducible representation are found as linear combinations of the original complete set of ionic displacements.

Although this machinery is elaborate for the problem at hand, it possesses some advantages:

(1) Very little actual calculation is involved if the representations of Kovalev[19] are employed.

(2) The role of time reversal is easily included, although in the present case this has trivial consequences.

(3) The coupling terms to strain and other macroscopic disturbances are easily derived.

(4) The approach is systematic, as opposed to a more intuitive method based on symmetries of polynomials.

(5) The calculation illustrates a method with wide application, easily adapted to other examples.

The use of multiplier representations has been discussed by Kovalev,[19] Lyubarski,[13] and by Maradudin

---

[19] O. V. Kovalev, *Irreducible Representations of the Space Groups* (Gordon and Breach, Science Publishers, Inc., New York, 1964).

and Vosko.[20] This latter treatment will be followed closely here. They have been applied to rutile by Gubanov and Shur.[21] An alternative to the multiplier representation approach is the factor group method, applied to rutile by Gay *et al.*[22]

V. COMPLETE SET OF IONIC DISPLACEMENTS

In the case of molecular vibrations, a complete set of displacements for a molecular structure of r ions can be constructed in a $3r$-dimensional space by taking as basis a set of $3r$ vectors each of which describes a unit displacement of one of the r ions along one of the three coordinate axes. For a crystal with r ions/unit cell the same set of $3r$ vectors can be defined for one unit cell. A set of displacements compatible with a given wave vector $\mathbf{q}$ is then defined for the whole crystal by, for example, propagating the jth of these $3r$ displacements to (say) the nth cell by a translation $\mathbf{x}(n)$ and multiplication by the phase factor $e^{i\mathbf{q}\cdot\mathbf{x}(n)}$. Denote the resulting displacement of the kth ion ($k=1, 2, \ldots, r$) in the nth cell by

$$e_{\mathbf{q}j}(k;i)e^{i\mathbf{q}\cdot\mathbf{x}(n)} \quad (j=1, 2, \ldots, 3r; i=1, 2, 3). \quad (8)$$

Any arbitrary displacement field in the rutile structure which transforms with wave vector $\mathbf{q}$ under translations of the rutile lattice can be expressed as a superposition of the $\{e_{\mathbf{q}j}\}$.

Paralleling the approach of Maradudin and Vosko,[20] define the general rutile space group operation $\{R/\mathbf{v}(R)+\mathbf{x}(n)\}$ by

$$\begin{aligned}\{R/\mathbf{v}(R)+\mathbf{x}(n)\}\cdot\mathbf{x}(m,k) &= R\mathbf{x}(m,k)+\mathbf{v}(R)+\mathbf{x}(n)\\ &\equiv \mathbf{x}(M,K),\end{aligned} \quad (9)$$

where R denotes rotation and reflection, $\mathbf{v}(R)$ is the associated nonprimitive translation, and $\mathbf{x}(n)$ is the general lattice translation of the rutile lattice. These various operations are considered to be referred to a common origin. This choice of origin affects the choice of $\mathbf{v}(R)$.

Define the $3r$-dimensional vector $\mathbf{E}_{\mathbf{q},j}$ by

$$E_{\mathbf{q},j}(s)=e_{\mathbf{q}j}(k;i), \quad [i=1, 2, 3; k=1, 2, \ldots, r; \; s=(k;i)=1, 2, \ldots, 3r]$$

where expression (8) defines the e's. It will be noted that although j and s both run through the same $3r$ values, they are not the same index, since j labels the $3r$ independent vectors and s labels components of these vectors. The $\mathbf{E}$'s transform according to the rule

$$\begin{aligned}\mathbf{E}_{\mathbf{q}j}' &\equiv \{R/\mathbf{v}(R)+\mathbf{x}(n)\}\mathbf{E}_{\mathbf{q}j}\\ &= \mathbf{\Gamma}[\mathbf{q}, \{R/\mathbf{v}(R)+\mathbf{x}(n)\}]\mathbf{E}_{\mathbf{q}j},\end{aligned} \quad (10)$$

where the $3r\times3r$ components of the matrix $\mathbf{\Gamma}$ are given by

$$\begin{aligned}&\Gamma_{ss'}[\mathbf{q}; \{R/\mathbf{v}(R)+\mathbf{x}(n)\}]\\ &\quad = R_{ii'}\delta(k,F(k',R))e^{i\mathbf{q}\cdot[\{R/\mathbf{v}(R)+\mathbf{x}(n)\}\cdot\mathbf{x}(k)-\mathbf{x}(k')]},\\ &\qquad (s=1, 2, \ldots, 3r; i=1, 2, 3; k=1, \ldots, r)\end{aligned} \quad (11)$$

with $F(k,R)$ defined through Eq. (9) by

$$K=F(k,R).$$

For $\mathbf{q}=0$, the matrices $\mathbf{\Gamma}$ are the direct product of the 3×3 fixed-point operation matrices $R_{ii'}$ and the $r\times r$ permutation matrices $\delta(k,F(k',R))$ just as would be the case in a vibration analysis of a single unit cell treated as if it were a molecule.

The matrices $\mathbf{\Gamma}$ commute with the $\mathbf{q}$th component of the Fourier-transformed dynamical matrix whose eigenvectors are the phonon eigenvectors for wavevector $\mathbf{q}$. The only symmetry operation of the Fourier transformed dynamical matrix which is not in the set of $\mathbf{\Gamma}$ matrices is the time reversal operation (see Maradudin and Vosko[20]). This point is very important because it identifies the symmetry analysis made here with the symmetry analysis of the crystal phonon spectrum, except for the effects of time-reversal symmetry. In the present case, it will be found that time-reversal symmetry has the effect only of allowing *real* phonon eigenvectors to be chosen at the ⟨101⟩ zone edge [see Eq. (15), *et seq.*].

VI. ⟨101⟩ BASIS VECTORS

One might hope to find exactly which phonons could be involved in the transition if each ⟨101⟩ phonon had a different symmetry, because presumably the observed displacements are compatible only with some of the symmetry types. Such an analysis has been performed for perovskite materials by Cochran and Zia.[23] Unfortunately, there are only two symmetry types at the ⟨101⟩ zone edge. As a result, it is possible only to say whether the phonons involved belong to one or both of these symmetry types. Since each of these symmetries appears more than once, even if only one type is unstable it cannot be decided by symmetry alone how many phonons of this type participate in the transition.

If several phonons of the same symmetry participate in the transition each one will contribute an identical series of terms to the free energy expansion. In addition there will be anharmonic terms coupling these phonons to each other. The number of coefficients available in such circumstances becomes too large to be useful, particularly since all the phonons will couple similarly to an external probe. However, by applying the minimization conditions for the free-energy expansion with respect to all the coefficients contributed by all but one of these modes, a one-phonon expression results in which the coefficients have a pressure and temperature de-

---

[20] A. A. Maradudin and S. H. Vosko, Rev. Mod. Phys. **40**, 1 (1968).

[21] A. I. Gubanov and M. S. Shur, Fiz. Tverd. Tela **7**, 2626 (1965) [English trans.: Soviet Phys.—Solid State **7**, 2124 (1966)].

[22] J. G. Gay *et al.*, J. Phys. Chem. Solids **29**, 1449 (1968).

[23] W. Cochran and A. Zia, Phys. Status Solidi **25**, 273 (1968).

TABLE I. Symmetry operations of D_{4h}^{14}.

Operation	Transformed position	Vanadium mixing	Oxygen mixing
$(E/0)$	(x,y,z)	(1)(2)	(3)(4)(5)(6)
$(C_{2x}/110)$	$(\frac{1}{2}+x,\ \frac{1}{2}-y,\ -z)$	(1,2)	(3,6)(4,5)
$(C_{2y}/110)$	$(\frac{1}{2}-x,\ \frac{1}{2}+y,\ -z)$	(1,2)	(3,5)(4,6)
$(C_{2a}/111)$	$(\frac{1}{2}+y,\ \frac{1}{2}+x,\ \frac{1}{2}-z)$	(1)(2)	(3)(4)(5,6)
$(C_{2b}/111)$	$(\frac{1}{2}-y,\ \frac{1}{2}-x,\ \frac{1}{2}-z)$	(1)(2)	(3,4)(5)(6)
$(C_4{}^2/0)$	$(-x,\ -y,\ z)$	(1)(2)	(3,4)(5,6)
$(C_4/001)$	$(-y,\ x,\ \frac{1}{2}+z)$	(1,2)	(3,6,4,5)
$(C_4{}^{-1}/001)$	$(y,\ -x,\ \frac{1}{2}+z)$	(1,2)	(3,5,4,6)
$(I/001)$	$(-x,\ -y,\ -z)$	(1)(2)	(3,4)(5,6)
$(\sigma_{vx}/11\bar{1})$	$(\frac{1}{2}-x,\ \frac{1}{2}-y,\ -\frac{1}{2}+z)$	(1,2)	(3,5)(4,6)
$(\sigma_{vy}/11\bar{1})$	$(\frac{1}{2}+x,\ \frac{1}{2}-y,\ -\frac{1}{2}+z)$	(1,2)	(3,6)(4,5)
$(\sigma_{va}/110)$	$(\frac{1}{2}-y,\ \frac{1}{2}-x,\ z)$	(1)(2)	(3,4)(5)(6)
$(\sigma_{vb}/110)$	$(\frac{1}{2}+y,\ \frac{1}{2}+x,\ z)$	(1)(2)	(3)(4)(5,6)
$(\sigma_h/001)$	$(x,\ y,\ \frac{1}{2}-z)$	(1)(2)	(3)(4)(5)(6)
$(S_4/0)$	$(y,\ -x,\ -z)$	(1,2)	(3,6,4,5)
$(S_4{}^{-1}/0)$	$(-y,\ x,\ -z)$	(1,2)	(3,5,4,6)

pendence which incorporates the anharmonic coupling effects to other phonons of the same symmetry. The result is the same as though only one phonon were treated from the outset, but of course, the physical interpretation is more complicated—in effect the burden of separating out the effect of phonon-phonon coupling has been removed from the phenomenological stage of determining the coefficients and thrown in with the problems associated with the determination of a microscopic model which must calculate the coefficients from first principles. One major effect of the presence of several phonons is that the distortion to which the crystal becomes unstable not only changes *amplitude* as the transition temperature is approached, but also changes direction (in the $3r$-dimensional space of rutile ion displacements) because the amplitude of the contribution of each participating phonon is temperature dependent. Such an effect must be explained by, e.g., a microscopic lattice dynamical model,[24] but it does not affect the form of the free-energy expansion. Phenomenologically then, only one phonon need be treated here.

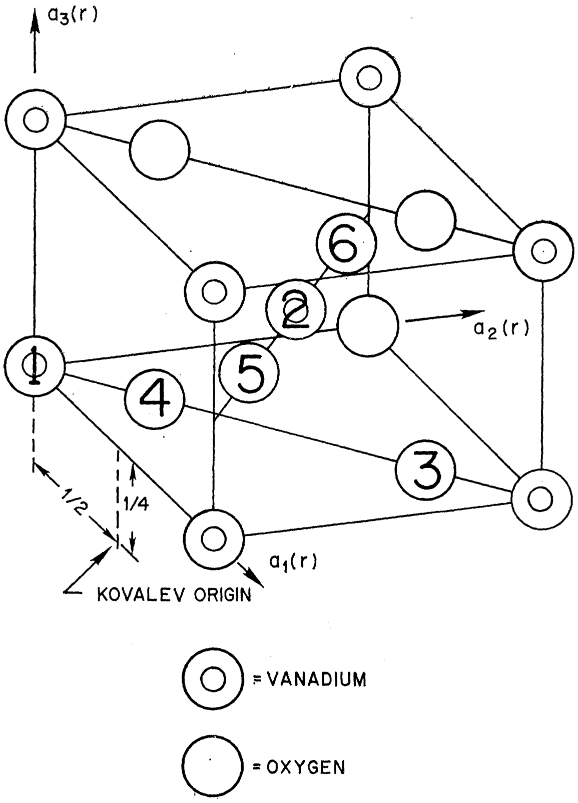

FIG. 1. Ionic labeling of the rutile phase of vanadium dioxide.

TABLE II. Ionic positions in rutile-phase vanadium dioxide. The position (a,b,c) is $[a\mathbf{a}_1(r)+b\mathbf{a}_2(r)+c\mathbf{a}_3(r)]$ relative to the origin of Kovalev. The parameter $x=0.305$.

k	Position $\mathbf{x}(k)$
1	$(-\frac{1}{2},\ 0,\ \frac{1}{4})$
2	$(0,\frac{1}{2},\frac{3}{4})$
3	$(-x-\frac{1}{2},\ -x,\ \frac{1}{4})$
4	$(x-\frac{1}{2},\ x,\ \frac{1}{4})$
5	$(x,\ \frac{1}{2}-x,\ \frac{3}{4})$
6	$(-x,\ \frac{1}{2}+x,\ \frac{3}{4})$

To proceed with the analysis, the six atoms of the rutile unit cell are labeled by the index k, as shown in Fig. 1. Under the space-group operations the ions are interchanged in the manner indicated in Table I.

Attention is now restricted to those space group elements in the group of the wave vector $\frac{1}{2}(\mathbf{b}_1+\mathbf{b}_3)$. Once a representation is constructed for this part of the space group, a representation for the full space group can also be generated (see Hamermesh).[25] This will be done in Sec. VII. Following Maradudin and Vosko[20] the matrices $T_{\alpha\beta}(k,k';\mathbf{q};R)$ are introduced:

$$T_{\alpha\beta}(k,k';\mathbf{q};R)=R_{\alpha\beta}\delta(k,F(k';R))e^{i\mathbf{q}\cdot[\mathbf{x}(k)-R\mathbf{x}(k')]} \qquad (12)$$

with

$$\mathbf{q}=\tfrac{1}{2}(\mathbf{b}_1+\mathbf{b}_3).$$

$F(k';R)$ is defined following Eq. (11).

The phases in Eq. (12) can be determined once the equilibrium positions $\mathbf{x}(k)$ of the ions in the rutile phase are known. These are listed relative to the origin of Kovalev[19] in Table II. The $\mathbf{T}$ matrices constitute a multiplier or ray representation of the point group of the wave vector D_{2h}, and satisfy the multiplication rule[20]:

$$\mathbf{T}(\mathbf{q};R_i)\cdot\mathbf{T}(\mathbf{q};R_j)=e^{i(\mathbf{q}-R_i^{-1}\mathbf{q})\cdot\mathbf{v}(R_j)}\cdot\mathbf{T}(\mathbf{q};R_i\cdot R_j), \qquad (13)$$

where the T's are the ray representation matrices for the group of the wave vector $\mathbf{q}$ and $\mathbf{v}(R_j)$ is the nonprimitive

---

[24] R. A. Cowley, Phil. Mag. **11**, 673 (1965).

[25] M. Hamermesh, *Group Theory and Its Application to Physical Problems* (Addison-Wesley Publishing Co., Inc., Reading, Mass., 1962), p. 458.

TABLE III. The two representations at the ⟨101⟩ zone edge of rutile. Also indicated are the characters of the matrices $T(R)$ discussed in the text, and some information used in determining the effect of time-reversal symmetry.

R	E	C_{2x}	C_{2y}	$C_4{}^2$	I	σ_{vx}	σ_{vy}	σ_h
$\tau^{(1)}$	$\begin{pmatrix}1&0\\0&1\end{pmatrix}$	$\begin{pmatrix}0&1\\1&0\end{pmatrix}$	$\begin{pmatrix}i&0\\0&-i\end{pmatrix}$	$\begin{pmatrix}0&-i\\i&0\end{pmatrix}$	$\begin{pmatrix}i&0\\0&i\end{pmatrix}$	$\begin{pmatrix}0&-i\\-i&0\end{pmatrix}$	$\begin{pmatrix}1&0\\0&-1\end{pmatrix}$	$\begin{pmatrix}0&1\\-1&0\end{pmatrix}$
$\tau^{(2)}$	$\begin{pmatrix}1&0\\0&1\end{pmatrix}$	$\begin{pmatrix}0&1\\1&0\end{pmatrix}$	$\begin{pmatrix}i&0\\0&-i\end{pmatrix}$	$\begin{pmatrix}0&-i\\i&0\end{pmatrix}$	$\begin{pmatrix}-i&0\\0&-i\end{pmatrix}$	$\begin{pmatrix}0&i\\i&0\end{pmatrix}$	$\begin{pmatrix}-1&0\\0&1\end{pmatrix}$	$\begin{pmatrix}0&-1\\1&0\end{pmatrix}$
$\chi_{T(R)}$	18	0	0	0	$6i$	0	0	0
Phase	1	-1	-1	1	-1	1	1	-1
$[\tau^{(1)}(R)]^2$	E	E	$-E$	E	$-E$	$-E$	E	$-E$

translation associated with the space-group operation $\{R_j/\mathbf{v}(R_j)\}$. These ray representation matrices are related to the matrices $\mathbf{\Gamma}$ of the true representation of Eq. (11) by

$$\mathbf{T}(\mathbf{q};R)=e^{i\mathbf{q}\cdot\mathbf{v}(R)}\mathbf{\Gamma}[\mathbf{q};\{R/\mathbf{v}(R)\}]. \tag{14}$$

The use of a ray representation enables one to restrict attention to the fixed-point operations R_i which are finite in number. This is done at the expense of introducing the phases of Eq. (13). Kovalev[19] has found that at the ⟨101⟩ zone edge for the rutile structure there are only the two irreducible representations $\tau^{(1)}$, $\tau^{(2)}$ listed in Table III. [The entries for x and y are interchanged in Kovalev, who treats $\mathbf{k}=(011)$ rather than (101).] The characters of the T matrices are also listed in Table III. It can be seen that $T(R)=6\tau^{(1)}(R)+3\tau^{(2)}(R)$, in agreement with Gubanov and Shur.[21]

That part of the $\mathbf{T}$ matrices based upon vanadium ions alone forms a representation by itself, since no mixing of atoms of different chemical species occurs under any symmetry operation. This vanadium ion representation is found to reduce to $3\tau^{(1)}$. Similarly, the oxygen ions form a representation reducible to $3\tau^{(1)}+3\tau^{(2)}$. The observed oxygen ion displacements of Andersson[5] are about one significant figure less accurate than the vanadium ion displacements. As a result the errors in the oxygen ion displacements are of the same order as the displacements themselves and so it is impossible to say whether or not there is a soft $\tau^{(2)}$ mode. The vanadium ion displacements are more accurate but because they form a representation reducible to $3\tau^{(1)}$ arbitrary displacements of the two vandium ions in the rutile unit cell are compatible with symmetry $\tau^{(1)}$.

Time reversal symmetry decides whether additional degeneracy of the phonon spectrum occurs. According to Maradudin and Vosko[26] no additional degeneracy occurs and, moreover, real phonon eigenvectors of symmetry $\tau^{(1)}$ can be chosen if

$$\frac{1}{h}\sum e^{-i(\mathbf{b}_1+\mathbf{b}_3)\cdot\mathbf{v}(R)}\chi[\tau^{(1)}(R)^2]=1, \tag{15}$$

where $\mathbf{v}(R)$ is the nonprimitive translation of the space group operation associated wlth the fixed point operation R, h is the order of the point group of wave vector ($h=0.8$ here), and χ is the trace of $\tau^{(1)}(R)\cdot\tau^{(1)}(R)$. The phase factors and matrices $[\tau^{(1)}(R)]^2$ are listed in Table III. Condition (15) can be seen to be satisfied. Using projection operators for the representation $\beta_1\tau^{(1)}(R)\beta_1{}^+$ with

$$\beta_1=\frac{1}{\sqrt{2}}\begin{pmatrix}i&1\\-i&1\end{pmatrix}$$

(chosen so that the representation matrices corresponding to $\mathbf{T}$-matrices which commute with the complex conjugation operation are real and, so far as possible, diagonal[20]) the eigenvectors for the vanadium

TABLE IV. The true representation matrices for the ⟨101⟩ zone edge. The real matrices in the second row are related to those in the first row by the unitary transformation $\tau_{(101)}{}^{(1)}(\text{real})=\beta\tau_{(101)}{}^{(1)}\beta^+$ with

$$\beta=\frac{1}{\sqrt{2}}\begin{pmatrix}1&-i\\1&i\end{pmatrix}.$$

	$(E/0)$	$(C_{2x}/110)$	$(C_{2y}/110)$	$(C_4{}^2/0)$	$(I/001)$	$(\sigma_{vx}/11\bar{1})$	$(\sigma_{vy}/11\bar{1})$	$(\sigma_h/001)$
$\tau_{(101)}{}^{(1)}$	$\begin{pmatrix}1&0\\0&1\end{pmatrix}$	$\begin{pmatrix}0&-i\\-i&0\end{pmatrix}$	$\begin{pmatrix}1&0\\0&-1\end{pmatrix}$	$\begin{pmatrix}0&-i\\i&0\end{pmatrix}$	$\begin{pmatrix}1&0\\0&1\end{pmatrix}$	$\begin{pmatrix}0&-i\\-i&0\end{pmatrix}$	$\begin{pmatrix}1&0\\0&-1\end{pmatrix}$	$\begin{pmatrix}0&-i\\i&0\end{pmatrix}$
$\tau_{(101)}{}^{(1)}(\text{real})$	$\begin{pmatrix}1&0\\0&1\end{pmatrix}$	$\begin{pmatrix}0&-1\\1&0\end{pmatrix}$	$\begin{pmatrix}0&1\\1&0\end{pmatrix}$	$\begin{pmatrix}1&0\\0&-1\end{pmatrix}$	$\begin{pmatrix}1&0\\0&1\end{pmatrix}$	$\begin{pmatrix}0&-1\\1&0\end{pmatrix}$	$\begin{pmatrix}0&1\\1&0\end{pmatrix}$	$\begin{pmatrix}1&0\\0&-1\end{pmatrix}$

---

[26] Reference 20, comment preceding Eq. (4.50a), p. 18, and Eq. (5.63), p. 28.

TABLE V. Full space group matrices generated by the basis $[f_1,f_2,f_3,f_4]$. All entries in these matrices are either zero or ±1. The unit entries are tabulated with their signs.

$(E/0)$	$(C_4/001)$	$(C_4^{-1}/001)$	$(C_4^2/0)$	$(C_{2a}/111)$	$(C_{2b}/111)$	$(C_{2x}/110)$	$(C_{2y}/110)$	$(I/001)$
11	13	13	11	−14	14	−12	12	11
22	24	−24	−22	23	23	21	21	22
33	−31	−31	−33	−32	32	−34	34	−33
44	42	−42	44	41	41	−43	−43	−44

ion representation are found to be

$$u_{(101),1}=\begin{bmatrix} a\\ b\\ 0\\ 0\\ 0\\ c\end{bmatrix},\quad u_{(101),2}=\begin{bmatrix} 0\\ 0\\ c\\ a\\ -b\\ 0\end{bmatrix},\qquad (16)$$

where the label (101) refers to displacements compatible with Eqs. (4) and (6). The first three entries are the x, y, and z components of the displacement of the vanadium ion with $k=1$; the second three entries are for $k=2$. Here a, b, c are real, and it is clear that by suitable choices of a, b, c three independent $u_{(101),1}$ can be constructed which, with their orthogonal partners $u_{(101),2}$, span the vanadium ion six-space. The difference of the two degenerate partners of Eq. (16), $u_{(101),1}-u_{(101),2}$, would describe the observed vanadium ion displacements if $a=b=-0.025a_1(r)$ and $c=-0.041a_3(r)$. With present knowledge of the phonon spectrum, it is impossible to say whether this is a rutile phonon eigenvector.

One might speculate that two modes are involved here, a "ferroelectric" mode and a "pairing" mode with eigenvectors,

$$\begin{bmatrix} -1\\ -1\\ 0\\ 1\\ -1\\ 0\end{bmatrix} \quad\text{and}\quad \begin{bmatrix} 0\\ 0\\ 1\\ 0\\ 0\\ -1\end{bmatrix},$$

respectively. The first describes a vanadium displacement inside its own oxygen cage similar to the titanium displacement which occurs in the cubic-tetragonal transition of $BaTiO_3$. The pairing mode involves only d-d overlap and might be considered to be a low-frequency mode on that account. Since even a small shift in frequency of a low-frequency mode is enough to drive it to zero frequency, a softening of the ferroelectric mode plus interaction with the pairing mode might drive the pairing mode to zero frequency. The hybridization of the two modes due to their interaction would cause this zero-frequency mode to share both a ferroelectric and a pairing character.

The conclusion of this first part of the symmetry analysis is that the observed distorted phase is compatible with one or more than one "frozen in" ⟨101⟩ phonon of symmetry $\tau^{(1)}$. Evaluation of the force constants of the crystal would allow an exact determination of the rutile phonon eigenvectors and hence a more definite decision as to the number of phonons participating in the transition, but for a phenomenological free-energy expansion this point need not be decided.

VII. REPRESENTATION OF FULL SPACE GROUP

It has been determined that if one or more ⟨101⟩ phonons of symmetry $\tau^{(1)}$ go soft it may be possible to explain the monoclinic distortion of VO_2. The representation $\tau^{(1)}$ is a ray representation of the group of the wave vector $\frac{1}{2}(\mathbf{b}_1+\mathbf{b}_3)$. The full space group contains the symmetry operation $(C_4/001)$ which maps this wave vector into $\frac{1}{2}(\mathbf{b}_2+\mathbf{b}_3)$. As a result (see, e.g., Lyubarski[27] or Hamermesh[25]) functions transforming according to this wave vector under the translation subgroup also occur among the independent basis functions for the irreducible representation of the full space group. Let the partners for wave vector $\frac{1}{2}(\mathbf{b}_1+\mathbf{b}_3)$ be $\{f_1,f_2\}$. Then the partners $\{f_3,f_4\}$ for wave vector $\frac{1}{2}(\mathbf{b}_2+\mathbf{b}_3)$ can be defined by

$$(C_4/001)f_j=f_{j+2}\quad (j=1,\,2).$$

The representation matrices for the full space group based upon all four $\{f_j\}$ are now fully determined.

First the ray representation $\tau^{(1)}$ is converted to a true representation of the group of the wave vector by multiplying by the phase factor $\exp\{-i(\mathbf{b}_1+\mathbf{b}_3)\cdot\mathbf{v}(R)/2\}$ [cf. Eq. (14)]. As a result the matrices in Table IV are obtained. Using the real representation matrices $\tau_{(101)}{}^{(1)}(\text{real})$ the basis vectors $\{f_j\}$ can also be chosen real. The full space group matrices generated by these real vectors are listed in Table V.

VIII. INVARIANTS OF $D_{4h}{}^{14}$

Using the matrices for the full group $D_{4h}{}^{14}$ just found, the representation generated by products such as $\{f_if_j\}$, $\{f_if_jf_k\}$, etc. are easily determined. These are then reduced and the combinations transforming as scalars under $D_{4h}{}^{14}$ are picked out.

For example, consider the representation $\mathbf{T}\times\mathbf{T}$ generated by $\{f_if_j\}$. Under pure translations the combinations $\{f_1f_1,f_1f_2,f_2f_2,f_3f_3,f_3f_4,f_4f_4\}$ transform according to $\mathbf{q}=0$ (modulo a reciprocal-lattice vector) and therefore reduce to a sum of the irreducible representations of the point group D_{4h}. On the other hand, the combination $\{f_1f_3,f_2f_3,f_1f_4,f_2f_4\}$ transform under pure transla-

---

[27] Reference 13, p. 91.

tions according to the representation $\mathbf{q}=\frac{1}{2}(\mathbf{b}_1+\mathbf{b}_2)$ and therefore reduce to a sum of irreducible ray representations of the group of this wave vector (i.e., ray representations of D_{4h}). There are no other possibilities.

A. Star $q=0$

This representation reduces into the representations $A_{1g}+A_{2g}+B_{1g}+B_{2g}+E_g$ of D_{4h}. The basis vectors are

$$\begin{aligned} A_{1g}&=[f_1^2+f_2^2+f_3^2+f_4^2],\\ A_{2g}&=[f_1^2-f_2^2+f_3^2-f_4^2],\\ B_{1g}&=[f_1^2+f_2^2-f_3^2-f_4^2],\\ B_{2g}&=[f_1^2-f_2^2-f_3^2+f_4^2],\\ E_g&=(f_1f_2,f_3f_4). \end{aligned} \tag{17}$$

B. Star $q=\frac{1}{2}(\boldsymbol{b}_1+\boldsymbol{b}_2)$

This representation is reducible to $\hat{\tau}_{10}+\hat{\tau}_4+\hat{\tau}_6$. The corresponding basis functions are

$$\begin{aligned} \hat{\tau}_{10}&=(f_1f_3,f_2f_4),\\ \hat{\tau}_4&=f_2f_3-if_1f_4,\\ \hat{\tau}_6&=f_2f_3+if_1f_4. \end{aligned} \tag{18}$$

From the above it follows that only one second-order invariant exists, that transforming as A_{1g} in Eq. (17). There is no third-order invariant, as can be seen from translation symmetry operations alone. Attention is therefore directed to fourth-order invariants.

IX. FOURTH-ORDER FREE-ENERGY TERM

To form the general fourth-order invariant, the shell theorem is employed. This states that if $\{f_j\}$ and $\{g_j\}$ both form a basis for the same representation of a group, then $\sum_j f_j^* g_j$ is an invariant. Imposing the requirement $\sum_{j=1}^4 f_j^2=\epsilon^2$ and letting

$$\begin{aligned} w_1&=1-2(f_2/\epsilon)^2-2(f_4/\epsilon)^2,\\ w_2&=1-2(f_2/\epsilon)^2-2(f_3/\epsilon)^2,\\ w_3&=1-2(f_3/\epsilon)^2-2(f_4/\epsilon)^2, \end{aligned} \tag{19}$$

the result is

$$F_4=\epsilon^4[A_0+A_1w_1^2+A_2w_2^2+A_3w_3^2]. \tag{20}$$

The requirement that the $(f_j/\epsilon)^2$ be positive and have unit sum restricts the $\{w_j\}$ to lie within a regular tetrahedron in w space with corners at (111), $(\bar{1}1\bar{1})$, $(1\bar{1}\bar{1})$, and $(\bar{1}1\bar{1})$.[28]

If it is now assumed that the fourth-order term in the free energy dominates higher-order terms, and if it is noted that the second-order term does not depend upon the $\{w_i\}$, then a minimum of the total free energy with respect to the $\{w_i\}$ is obtained by extremizing the fourth-order term in w space. Since a first-order transition is observed, the fourth-order term is expected to be negative[7,12] and so a minimum in the free energy will be obtained by *maximizing* the fourth-order term with respect to the $\{w_i\}$. Suppose

$$A_1, A_2<0,$$
$$A_3>0.$$

(Other inequalities lead to space groups other than $C_{2h}{}^5$.) Then F_4 is maximized if

$$\begin{aligned} w_1&=w_2=0,\\ |w_3|&=1. \end{aligned} \tag{21}$$

One solution compatible with (21) is

$$\begin{aligned} f_1^2&=f_2^2=\tfrac{1}{2}\epsilon^2,\\ f_3^2&=f_4^2=0, \end{aligned} \tag{22}$$

which leads to a displacement field

$$u_{(101)}=c[f_1\pm f_2].$$

Taking the positive sign and referring to the matrices of Table IV, this ground state is invariant under the space-group operations $\{E/0\}$, $\{C_{2y}/110\}$, $\{I/001\}$, $\{\sigma_{vy}/11\bar{1}\}$ which are shown to be symmetry operations of $C_{2h}{}^5$ in the Appendix. It is also invariant under the translations

$$\mathbf{x}(m)=m_3\mathbf{a}_1(r)+m_2\mathbf{a}_2(r)+(2m_1-m_3)\mathbf{a}_3(r)$$

of the monoclinic phase of vanadium dioxide.

Another set of solutions compatible with (21) is

$$\begin{aligned} f_1^2&=f_2^2=0,\\ f_3^2&=f_4^2=\tfrac{1}{2}\epsilon^2, \end{aligned} \tag{23}$$

leading to

$$u_{(011)}=C[f_3\pm f_4].$$

Taking the negative signs the ground state is invariant under the translations

$$\mathbf{x}(m)=m_3\mathbf{a}_1(r)+m_2\mathbf{a}_2(r)+(2m_1-m_2)\mathbf{a}_3(r)$$

and also the operations $\{E/0\}$, $\{C_{2x}/110\}$, $\{I/003\}$, and $\{\sigma_{vx}/11\bar{3}\}$ of the space group $C_{2h}{}^5$: $P2_1/C$. This is just a $\frac{1}{2}\pi$ rotation of the first solution. Isomorphic possibilities result from the other choices of sign.

X. $q=0$ STAR AND INCLUSION OF STRAIN

It was noted earlier that besides the ⟨101⟩ star there was allowed the star $\mathbf{q}=0$. The straightforward procedure would be to reduce the $\mathbf{T}$-matrix representation at $\mathbf{q}=0$ and form the linear, quadratic, etc., invariants as was just done for the zone edge. However, such a procedure is inadequate to include the effect of strain on the crystal, even though the strain is expressible in terms of zone-center acoustic modes. This difficulty arises because the strain is described only in terms of a limit of the acoustic mode amplitudes as $\mathbf{q}$ tends to zero, and this limit is not well-defined but depends upon the direction along which $\mathbf{q}$ is allowed to tend to zero

---

[28] The author is indebted to E. I. Blount for this observation.

(see, e.g., Born and Huang[29]). It is therefore necessary to consider the acoustic modes for infinitesimal $\mathbf{q}$ to properly include the strain effects. Rather than dealing with these modes and then reinterpreting the results in terms of strain (cf. Kwok and Miller[8]) the procedure followed here will be to treat the strain parameters, S_{ij}, themselves from the outset. Strain parameters have been used as order parameters previously by Anderson and Blount[30] and by Boccara.[31]

The $\mathbf{q}=0$ limit for modes other than the acoustic modes is unphysical in two ways.[32] First, for extremely small $\mathbf{q}$ the finite sample shape causes a departure from the symmetry considerations of the infinite sample considered here. Secondly, for somewhat larger $\mathbf{q}(\approx\omega/c)$ polariton effects due to coupling of the lattice vibrations with the transverse electromagnetic field become important. Since our symmetry arguments are based upon an *electrostatic* model they do not necessarily apply in the polariton regime. Rigor therefore demands that our symmetry arguments be made for nonzero $\mathbf{q}$. Such an analysis would differ from the $\mathbf{q}=0$ analysis primarily in predicting a splitting of the LO and TO modes, as is observed in all polar materials. A $\mathbf{q}=0$ analysis is therefore based on the premises that the most important couplings will be those allowed even under conditions of highest degeneracy, and also that this strong coupling will extend over some range of $\mathbf{q}$'s, not merely exactly at zone center. With these reservations we will make a $\mathbf{q}=0$ analysis. The reduction of the $\mathbf{q}=0$ $\mathbf{T}$-matrix representation is

$$A_{1g}+A_{2g}+A_{2u}+B_{1g}+2B_{1u}+B_{2g}+E_g+3E_u\,,$$

where the acoustic modes have been removed. Since the $\mathbf{q}=0$ representation based upon $\{f_if_j\}$ products reduces according to Eq. (17) into

$$A_{1g}+A_{2g}+B_{1g}+B_{2g}+E_g$$

it is clear that the $\langle 101\rangle$ modes can couple linearly with all the $\mathbf{q}=0$ Raman active modes and also with the silent mode A_{2g}. However, no coupling linear in the amplitudes of the infrared active modes occurs, and hence coupling to macroscopic polarization must be quadratic or higher in the polarization (see Sec. XI). For reasons of simplicity, only the strain and strain coupling terms will be included from the $q=0$ analysis.

The free energy to fourth order is found to be

$$\begin{aligned}F=F_0&+\alpha(S_{11}+S_{22})+\beta S_{33}+C_{11}(S_{11}^2+S_{22}^2)+C_{12}S_{11}S_{22}\\&+C_{13}(S_{11}+S_{22})S_{33}+C_{33}S_{33}^2+C_{44}(S_{23}^2+S_{31}^2)\\&+C_{66}S_{12}^2+\epsilon^2F_2+\epsilon^2[(B_0+B_1w_3)S_{11}+(B_0-B_1w_3)S_{22}\\&+B_2S_{33}+B_3S_{12}w_2]+B_4(S_{31}f_1f_2+S_{23}f_3f_4)\\&+\epsilon^4[A_0+A_1w_1^2+A_2w_2^2+A_3w_3^2].\end{aligned}\tag{24}$$

Under the conditions of zero external stress the derivatives of the free energy with respect to the strains must vanish. Imposing these conditions, the strains naturally associated with zone edge distortion are found to be

$$(S_{11}+S_{22})=-[2C_{33}[\alpha+B_0\epsilon^2]-[\beta+B_2\epsilon^2]C_{13}]/\Delta\,,$$

$$S_{33}=-[(C_{11}+\tfrac{1}{2}C_{12})[\beta+B_2\epsilon^2]-C_{13}(\alpha+B_0\epsilon^2)]/\Delta\,,$$

where

$$\Delta=C_{33}(2C_{11}+C_{12})-C_{13}^2$$

and

$$\begin{aligned}(S_{11}-S_{22})&=2B_1w_3\epsilon^2/(2C_{11}-C_{12})\,,\\S_{23}&=-B_4f_3f_4/(2C_{44})\,,\\S_{31}&=B_4f_1f_2/(2C_{44})\,,\\S_{12}&=-B_3w_2\epsilon^2/(2C_{66})\,.\end{aligned}\tag{25}$$

The coefficients α and β are related to thermal expansion coefficients. The $\{C_{ij}\}$ are the rutile elastic constants.

From Eq. (25) it follows that the strains are second order in the distortion, apart from the first two strains, $S_{11}+S_{22}$ and S_{33}, which represent volume changes with no accompanying change of symmetry. Since the transition to a state of space group C_{2h}^5 has been found to require $w_1=w_2=0$ and $|w_3|=1$ [see Eq. (21)], $S_{11}-S_{22}\neq 0$, $S_{12}=0$ and one of (S_{13},S_{23}) is zero.

In Sec. III and again following Eq. (23) it was pointed out that two of the degenerate ground states available in the low-temperature phase are related by a fourfold screw axis parallel to the rutile c axis. If a macroscopic sample of VO_2 forms domains simply by one region choosing one of these ground states and another region a screw-related ground state, the strain $S_{11}-S_{22}$ will be of one sign for the domains obeying Eq. (22) and of the opposite sign for those obeying Eq. (23). That is, in both domain types one a-axis dimension of the rutile unit cell will be contracted relative to the other, but each domain makes a different choice of which a-axis dimension to shorten. Domains related by a ninety degree rotation about the c axis have been observed.[33] This may be the cause of VO_2 sample cracking along the rutile c axis observed[6] when the samples are cooled through the transition temperature.

If the various strains of Eq. (25) are now substituted back into the free-energy expression of Eq. (24) and the terms in $\{w_i^2\}$ grouped together the result is again the expression of Eq. (20) but with new coefficients which now involve the various coupling constants $\{B_i\}$ as well as the original $\{A_i\}$. Denoting these modified coefficients by $\{A_i'\}$ the free-energy expansion is of the form

$$F=F_0+\epsilon^2F_2+\epsilon^4[A_0'+\sum_{i=1}^{3}A_i'w_i^2]+\cdots\,,\tag{26}$$

---

[29] M. Born and K. Huang, *Dynamical Theory of Crystal Lattices* (Oxford University Press, New York, 1954), p. 230 ff.

[30] P. W. Anderson and E. I. Blount, Phys. Rev. Letters **14**, 217 (1965).

[31] N. Boccara, Ann. Phys. (N. Y.) **47**, 40 (1968).

[32] S. S. Mitra, in *Optical Properties of Solids*, edited by S. Nudelman and S. S. Mitra (Plenum Press, Inc., New York, 1969), p. 362.

[33] A. S. Barker *et al.*, Phys. Rev. Letters **17**, 1286 (1966)

with

$$A_{1,2}'<0\,,$$
$$A_3'>0\,,$$

and $|w_i|\leq 1$ inside the regular tetrahedron described following Eq. (20).

On minimizing with respect to the $\{w_i\}$, Eq. (26) is found to take the form

$$F=F_0+\epsilon^2F_2+\epsilon^4(A_0'+A_3')\,. \qquad (27)$$

If the entire fourth-order contribution to the free energy were that displayed above, including possible coupling to other effects like $\mathbf{q}=0$ phonons, the experimental fact that the transition is first order would imply that

$$A_0'<-A_3'. \qquad (28)$$

This also means that powers of ϵ higher than the fourth must be examined to determine the spontaneously adopted value of ϵ and hence the temperature dependence of ϵ below the transition temperature.

XI. COUPLING TO ELECTRIC FIELD

If an electric field E_A is applied to the crystal a field

$$E=E_A+E_C$$

will appear in the crystal, where E_C is the depolarizing field due to polarization of the sample. The crystal energy will then be changed by additional terms involving the dielectric constant of the crystal. If the field is not so strong that morphic effects must be considered (i.e., the lowering of crystal symmetry owing to distortion of the lattice in response to the field can be ignored), then by analogy with the strain contributions in Eq. (24) the presence of an electric field adds to the free energy the terms F_E,

$$\begin{aligned}F_E=\kappa_a(E_x^2+E_y^2)+\kappa_cE_z^2+\epsilon^2\{(D_0+D_1w_3)E_x^2\\ +(D_0-D_1w_3)E_y^2+D_2E_z^2+D_3w_2E_xE_y\}\\ +D_4(f_1f_2E_xE_z+f_3f_4E_yE_z)\,, \qquad (29)\end{aligned}$$

where κ_a, κ_c are the dielectric constants of the rutile phase normal to the c axis and along the c axis, respectively. The dielectric constant of the material for the phase of Eq. (22) is then

$$\left.\begin{aligned}\kappa_{11}&=\kappa_a+\epsilon^2(D_0+D_1)\,,\\ \kappa_{22}&=\kappa_a+\epsilon^2(D_0-D_1)\,,\\ \kappa_{33}&=\kappa_c+\epsilon^2D_2\,,\\ \kappa_{13}&=\epsilon^2D_4\,.\end{aligned}\right. \qquad (30)$$

Measurement of $\{\kappa_{ij}\}$ therefore provides a direct monitoring of ϵ^2 as a function of (say) temperature and volume. Unfortunately the high conductivity of even the semiconducting state due to various sample defects[6] makes such an experiment difficult.

XII. EXPERIMENTAL EVIDENCE

Evidence which could assist in an empirical determination of the coefficients in the free-energy expansion is not plentiful. To compare the experiments with the expansion of Eq. (24) we define two temperatures: T_t, the first-order transition temperature, and T_c, the temperature at which a second-order transition (not to $C_{2h}{}^5$) would occur if F_4 were positive.

Then, assuming that the second term can be rewritten as

$$F_2\epsilon^2=(T-T_c)a\epsilon^2(T)\,,$$

the measured volume change[34] and latent heat[6,35] provide the estimate

$$\left[1+(T_t-T_c)\frac{1}{\epsilon^2}\frac{d\epsilon^2}{dT_t}\right]a\epsilon^2(T_t) =L/T_t\sim 3 \text{ cal mole °K}. \qquad (31)$$

Measurements of the discontinuities in cell dimensions made by Minomura and Nagasaki[35] ($\Delta a_r=-\Delta b_r=-\Delta c_r=0.03$ Å) indicate that at T_t

$$\frac{B_1\epsilon^2}{2C_{11}-C_{12}}\simeq 10^{-2}\,,$$

$$\frac{\epsilon^2}{\Delta}|2C_{33}B_0-B_2C_{13}|\ll 10^{-2}\,,$$

$$[B_2(C_{11}+\tfrac{1}{2}C_{12})-B_0C_{13}]\epsilon^2/\Delta\simeq -10^{-2}.$$

The notation is that of Eq. (25). Including thermal strain present at T_t in F_0 and making a Taylor series expansion of α and β about $T=T_t$ the linear coefficients of expansion are found to be

$$\alpha_a=\alpha_b=-[C_{33}\alpha'(T_t)-\tfrac{1}{2}C_{13}\beta'(T_t)]/\Delta\,,$$
$$\alpha_c=-[(C_{11}+\tfrac{1}{2}C_{12})\beta'(T_t)-C_{13}\alpha'(T_t)]/\Delta\,,$$

where, e.g.,

$$\alpha'(T_t)=\left.\frac{d\alpha}{dT}\right|_{T=T_t}.$$

It is then found that the changes in the coefficients of expansion on passing through the transition are

$$\Delta\alpha_a\simeq-\Delta\alpha_b\simeq\left(\frac{\partial\epsilon^2}{\partial T_t}\right)_V B_1/(2C_{11}-C_{12})\,,$$

$$\Delta\alpha_c=[B_2(C_{11}+\tfrac{1}{2}C_{12})-B_0C_{13}]\left(\frac{\partial\epsilon^2}{\partial T_t}\right)_V\Big/\Delta.$$

The derivatives are evaluated at a temperature just below T_t. The change in the linear coefficient of ex-

---

[34] T. Kawakubo and T. Nakagawa, J. Phys. Soc. Japan **19**, 517 (1964).

[35] S. Minomura and H. Nagasaki, J. Phys. Soc. Japan **19**, 131 (1964).

pansion for a polycrystalline rod is then roughly

$$\Delta\alpha=\tfrac{1}{3}\Delta\alpha_c=-\tfrac{1}{3}\times10^{-2}\frac{1}{\epsilon^2}\left(\frac{\partial\epsilon^2}{\partial T_t}\right)_V .$$

Assuming the derivative to be negative since it is expected that ϵ^2 decreases with temperature near T_t the sign of the change agrees with Kawakubo *et al.*[34] and the estimate

$$\frac{1}{\epsilon^2}\left(\frac{\partial\epsilon^2}{\partial T_t}\right)_V=-3\times10^{-3}/°\mathrm{K}$$

is obtained. This means that the prefactor in Eq. (31) is probably near unity.

Additional information has been obtained by Ladd and Paul.[36] These authors have measured the change in transition temperature under hydrostatic pressure and under uniaxial stress along the c axis and the a axis. They find $(dT_t/dS_c)=-1.2\times10^{-3}$°K/bar along the c axis and $(dT_t/dP)=6\times10^{-5}$°K/bar for hydrostatic pressure. The transition temperature was so broadened for a-axis stress that no meaningful result could be quoted. The broadening of the transition temperature under a-axis or b-axis stress which they observe can be attributed to domain structure of the kind proposed in this paper to account for c-axis cracking. From the free energy expansion it is found that the leading term in the change of T_t under pressure is due to the shift in T_c, which is

$$(dT_c/dS_c)=B_3(C_{11}+\tfrac{1}{2}C_{12})/(\Delta a)\,,$$
$$(dT_c/dS_a)=(B_0+B_1w_3)2C_{33}/(\Delta a)\,,$$
$$(dT_c/dS_b)=(B_0-B_1w_3)2C_{33}/(\Delta a)\,.$$

That is, those domains for which $w_3=1$ have a larger shift in transition temperature under a-axis stress than those for which $w_3=-1$ (assuming B_0 and B_1 have the same sign; vice versa if the signs are opposite). As a result a smearing out of the transition temperature will occur for a multidomain sample.

XIII. SUMMARY AND REMARKS

In this paper the background of the symmetry analysis of the free-energy expansion has been laid out in some detail, particularly in its discussion of the role played by the various degenerate low-temperature ground states. The analysis also has been developed so as to emphasize that at no point must one assume that the observed distortion is expressible as one rutile-phase zero-frequency phonon, although if in fact the distortion is produced by a "frozen-in" phonon the analysis decides its symmetry. This caution has been taken because the transition is first order and also because the unstable phonon is zone edge, where many phonons of the same symmetry exist. Both these facts lead one to suspect that the instability may really involve a number of phonons, and that the phonon spectrum in the transition region of temperatures may depart considerably from that of the parent rutile phase.

It has been found that the distorted monoclinic phase can be described as a ⟨101⟩ zone-edge wave-vector displacement field. There are only two symmetries for displacements at this point on the zone edge, and there are six distinct doubly degenerate displacement fields of the correct symmetry, $\tau^{(1)}$, to describe the observed displacements. Because there are six of them, their form is not determined by symmetry alone. If one wished to choose the displacement fields to correspond to rutile-phase phonons, the dynamical matrix for the vanadium dioxide phonons would have to be known. It is not. Therefore it cannot be decided whether one or more than one phonon is involved in the monoclinic transformation. However, one could say that the experimentally determined vanadium displacements[5] are compatible with only one phonon going unstable. Whether or not this is the case, the free-energy expansion remains valid.

It has also been found that more accurate oxygen positions are needed for the monoclinic phase to rule out an unstable $\tau^{(2)}$ mode additional to the definite instability of one or more $\tau^{(1)}$ modes.

Following the determination of the various symmetry allowed eigenvectors at the ⟨101⟩ zone edge the free energy has been expanded to fourth order in these displacements. The strain terms also have been included in this expansion. It has been shown that a transition from $D_{4h}{}^{14}$ to $C_{2h}{}^{5}$ can be described by such an expansion and that the transition may be of first order.

Observations which bear upon the validity of the free-energy expansion model are not plentiful. One such observation is the cracking of $\mathrm{VO_2}$ samples along the rutile c-axis upon the transition to the low temperature phase. This cracking has been attributed in the present paper to the formation of domains related by a 90° screw axis parallel to the rutile c axis. Observations of these domains are explained by the model.[33,35] It has also been shown that the smearing out of the transition temperature under a-axis stress observed by Ladd and Paul[36] might be explained by such a domain structure.

Work on the dependence of the transition temperature on electronic repopulation effects[6] indicates that changes in the free energy due to the redistribution of those electrons near in energy to the band gap of the semiconducting phase are not major influences upon the coefficients. The striking resistivity change accompanying the transition which has stimulated the formation of numerous narrow-band microscropic models[1] of the transition can probably be explained[6] as the distortion-induced separation of two bands which overlap for $T>T_c$. Such a removal of overlap could lead to a semiconducting behavior for $T<T_c$. If this is the case the overlap change would be proportional to some power of ϵ.

---

[36] L. A. Ladd and W. Paul, Solid State Commun. **7**, 425 (1969).

The present analysis does not lead to a macroscopic interpretation for ϵ. However, ϵ does couple to a homogeneous electric field, though not so directly as in the case of ferroelectrics.[7] The analysis of this paper also shows a coupling to strain and to zone center Raman active phonons. Any of these effects might conceivably lead to a means of monitoring the distortion parameter ϵ as a function of temperature and pressure, the key experiment for determining the coefficients of the free-energy expansion derived in this paper.

ACKNOWLEDGMENTS

The author wishes to thank C. N. Berglund for his encouragement and for suggesting the problem. It is also a pleasure to acknowledge comments by R. M. Ryder and J. M. Worlock. He particularly thanks E. I. Blount for his useful advice.

APPENDIX: CRYSTAL SYMMETRY, D_{4h}^{14} AND C_{2h}^{5}

The symmetry operations for the space group of rutile (D_{4h}^{4}:$P4_2/mnm$) and for the monoclinic phase (C_{2h}^{5}:$P2_1/c$) will be denoted by the symbol $(R/a,b,c)$, where R is a fixed point operation (e.g., a reflection) and $\frac{1}{2}(a,b,c)$ are the components of a nonprimitive translation, viz., the translation $\frac{1}{2}a\mathbf{a}_1+\frac{1}{2}b\mathbf{a}_2+\frac{1}{2}c\mathbf{a}_3$ where the vectors $\{\mathbf{a}_i\}$ define the lattice. The factor $\frac{1}{2}$ is just a convenience.

A. Translations

The basis vectors for the rutile lattice are orthogonal and satisfy

$$|\mathbf{a}_1|=|\mathbf{a}_2|\,.$$

The general translation in the rutile structure is then

$$\mathbf{x}(r)=m_1\mathbf{a}_1+m_2\mathbf{a}_2+m_3\mathbf{a}_3 \quad (m_1,\, m_2,\, m_3 \text{ integral})\,. \tag{A1}$$

The basis vectors for the monoclinic phases are experimentally observed to satisfy[5]

$$\begin{aligned} \mathbf{a}_1(m)&=2\mathbf{a}_3(r)\,,\\ \mathbf{a}_2(m)&=-\mathbf{a}_2(r)\,,\\ \mathbf{a}_3(m)&=\mathbf{a}_1(r)-\mathbf{a}_3(r)\,, \end{aligned} \tag{A2}$$

where m denotes monoclinic and r denotes rutile. The general translation in the monoclinic structure is then

$$\mathbf{x}(m)=m_3\mathbf{a}_1(r)+m_2\mathbf{a}_2(r)+(2m_1-m_3)\mathbf{a}_3(r) \quad (m_1,\, m_2,\, m_3 \text{ integral})\,. \tag{A3}$$

B. Other Operations

The remaining operations involving rotations, reflections, screws, and glides can all be referred to a common origin by adjusting the associated translations. Standard coordinate systems for all the crystal structures have been set up in the *International Tables of X-Ray Crystallography*.[37] Referring to these standard coordinate systems Kovalev[19] adopts the origin $\frac{1}{2}\mathbf{a}_1(r)-\frac{1}{4}\mathbf{a}_3(r)$ for the rutile structure and $\frac{1}{4}\mathbf{a}_2(m)+\frac{1}{4}\mathbf{a}_3(m)$ for the monoclinic structure. The rutile symmetry operations (apart from translations) are then combinations of

$(C_{2x}/110)=$ twofold rotation about x axis, followed by a translation through $\frac{1}{2}(\mathbf{a}_1+\mathbf{a}_2)$,
$(C_{2y}/110)=$ same, but about y axis,
$(C_4^2/000)=$ twofold rotation about z axis,
$(C_{2a}/111)=$ twofold rotation about (110) followed by $\frac{1}{2}(\mathbf{a}_1+\mathbf{a}_2+\mathbf{a}_3)$,
$(C_{2b}/111)=$ same but about $(1\bar{1}0)$,
$(C_4/000)=\frac{1}{2}\pi$ rotation about z axis,
$(C_4^{-1}/000)=\frac{3}{2}\pi$ rotation about z axis,
$(I/001)=$ inversion followed by a translation of $\frac{1}{2}\mathbf{a}_3$.

For the monoclinic structure, C_{2h}^{5}, symmetry operations (apart from translations) are combinations of

$(C_4^2/001)=$ twofold rotation about $\mathbf{a}_3(K)$,
$(I/101)=$ inversion $+\frac{1}{2}[\mathbf{a}_1(K)+\mathbf{a}_3(K)]$,
$(\sigma_h/100)=$ reflection in plane perpendicular to $\mathbf{a}_1(K)$ followed by $\frac{1}{2}\mathbf{a}_1(K)$.

Here $\{\mathbf{a}_i(K)\}$ are the Kovalev basis vectors which are defined by

$$\begin{aligned} \mathbf{a}_1(K)&=\mathbf{a}_3(m)\,,\\ \mathbf{a}_2(K)&=\mathbf{a}_1(m)\,,\\ \mathbf{a}_3(K)&=\mathbf{a}_2(m)\,. \end{aligned}$$

Expressing the monoclinic symmetry operations in the rutile coordinate system and referring them to the rutile origin

$$(0,0,0)_m=(-\tfrac{1}{4},\,\tfrac{1}{4},\,\tfrac{1}{2})_r$$

these symmetry operations become

$$(C_{2y}/110)\,,\ (I/001)\,,\ (\sigma_{vy}/11\bar{1})\,,$$

where the translations are also in the rutile basis.

It follows from Subsec. 1 of this appendix that the monoclinic translation group is a subgroup of the rutile translation group, and from Subsec. 2 that the other space-group operations of C_{2h}^{5} are contained in D_{4h}^{14}. This justifies one basic assumption of this paper, viz., that the VO_2 phase transition is a case of *lowered* symmetry, not merely one of *changed* symmetry.

---

[37] K. Lonsdale, *International Tables for X-Ray Crystallography* (Kynoch Press, Birmingham, England, 1952), 3 Vols.