

Stopping Power of Swift Hydrogen and Helium Ions in Gases

By F. BESENBACHER, H. H. ANDERSEN,
P. HVELPLUND, *and* H. KNUDSEN

Det Kongelige Danske Videnskabernes Selskab
Matematisk-fysiske Meddelelser 40:3



Kommissionær: Munksgaard

1979

Contents

	Page
§1. Introduction	3
§2. Theory	3
§3. Experimental procedure and data treatment	7
§4. Experimental results and comparison with other data	15
§5. Comparison with theory	26
§6. Conclusion	37
References	38

Synopsis

Stopping powers of H_2 , He, N_2 , O_2 , CO_2 , Ne, Ar, Kr, and Xe for 40-keV to 1-MeV hydrogen ions and 100-keV to 2.4-MeV helium ions have been measured to an accuracy of $\pm 2.5\%$ (2σ). The stopping powers for hydrogen show good agreement with most other published results and with the Andersen and Ziegler tabulations, while those for helium ions are up to 6% lower than the helium data of Chu and Powers. With higher-order Z_1 correction terms included in the theoretical description, Bonderup's calculated shell-corrections based on the Lindhard-Scharff model, are in good agreement with the experimental proton data for $E_p \geq 100$ keV, and experimental I values may be deduced. Within the velocity region $4.4 \lesssim v/v_0 \lesssim 4.9$ and for target materials with $1 \leq Z_2 \leq 54$, the experimental findings support Lindhard's and Esbensen's value for the Barkas correction. The stopping-power ratios S_{He}/S_H depend strongly on Z_2 and deviate significantly from the mean-square charge state obtained from experimental equilibrium charge state distributions.

Institute of Physics, University of Aarhus
DK-8000 Aarhus C, Denmark

§1. Introduction

The energy loss of an ion beam traversing matter is a phenomenon of basic interest to atomic physics and has been the subject of much theoretical and experimental work. However, there is still a great need for a better understanding of the details of the stopping processes.

It is the purpose of the present work, through accurate measurements to test the Lindhard-Scharff model (Lindhard and Scharff 1953, 1960) and its refinements for the average energy-loss (Bonderup 1967) and energy straggling (Bonderup and Hvelplund 1971). Especially it is of interest to examine the so-called shell corrections at the present rather low energies, and to investigate Sigmund's molecular correlation effect in energy straggling (Sigmund 1976).

A detailed understanding of the underlying mechanisms is imperative for obtaining the kind of comprehensive and accurate energy loss and straggling information necessary in, for example, a composition analysis of thin films via Rutherford-backscattering or nuclear reaction techniques.

We have carried out a systematic investigation of the stopping power and energy straggling for hydrogen and helium ions in H_2 , He, N_2 , O_2 , CO_2 , Ne, Ar, Kr, and Xe at ion energies $40 \text{ keV} \lesssim E_H \lesssim 1 \text{ MeV}$ and $100 \text{ keV} \lesssim E_{He} \lesssim 2.4 \text{ MeV}$. Gaseous targets were chosen so as to avoid specific solid-state effects that might obscure especially the straggling results.

The present paper deals with the stopping-power results only. A forthcoming publication (Besenbacher et al. 1980) will deal with the straggling results, some of which have already been published (Besenbacher et al. 1977).

After a brief review of the energy-loss theory in §2, the experimental procedure and data treatment will be discussed in §3. In §4, the experimental results will be presented and compared with empirical stopping-power tabulations and other published results and, finally, in §5, the experimental data will be discussed and compared with theory.

§2. Theory

As discussed by Bohr (1948), two distinctly different mechanisms are responsible for the slowing-down of nonrelativistic charged particles: (i) Electronic stopping, i. e., energy loss to excitation and ionization, and (ii) Nuclear stopping, i. e., energy transfer leading to translatory motion of the struck atom as a whole. In the present velocity range $v \gtrsim v_0$ (v_0 is the Bohr velocity e^2/\hbar), and with our experimental geometrical arrangement, the nuclear energy loss is almost negligible (see below).

At high velocities, where

$$\kappa_B = \frac{2Z_1 v_o}{v} < 1, \quad (1)$$

a quantal perturbation treatment is applicable and gives the wellknown Bethe formula (Bethe 1930). According to this formula, the specific electronic energy loss suffered by a heavy incoming particle with charge $Z_1 e$ and velocity v penetrating a target of atomic number Z_2 and density N (atoms per unit volume) is given by

$$\left(-\frac{dE}{dR}\right)_e = \frac{4\pi Z_1^2 e^4}{mv^2} N Z_2 L(v, Z_2) \quad (2)$$

where

$$L(v, Z_2) = L_o = \log\left(\frac{2mv^2}{I}\right) - \log\left(1 - \left(\frac{v}{c}\right)^2\right) - \left(\frac{v}{c}\right)^2 - \frac{C(v, Z_2)}{Z_2} \quad (3)$$

Here, m and $-e$ are the electron mass and charge, respectively, and c the velocity of light. The main parameters of the Bethe formula are the mean ionization potential I and the shell corrections C/Z_2 . The former is defined by

$$\log I = \frac{1}{Z_2} \sum_i f_{oi} \log \hbar \omega_{oi} \quad (4)$$

where f_{oi} are the dipole oscillator strengths corresponding to the transition frequencies ω_{oi} for the atomic system. A direct calculation of I from this formula has until recently been impeded due to the lack of knowledge about the distribution of oscillator strength in the relevant energy region, i. e., from 10-1000 eV. In a Thomas-Fermi model the calculation of I is very much simplified since f_{oi} is a function of ω/Z_2 only and Bloch (1933) showed within this model that

$$I/Z_2 = I_o = \text{constant}. \quad (5)$$

This result is in qualitative agreement with the experimental results for heavier atoms and empirically I_o is of the order of 10 eV. For lighter atoms the cut-off in $f(\omega, Z_2)$ close to the Rydberg frequency leads to a somewhat higher value of I/Z_2 (Lindhard 1964).

To calculate $L(v, Z_2)$ in formula (2) and thereby to determine the mean ionization potential and the shell corrections, Lindhard and Scharff (1953, 1960) considered the target as a collection of free electron gases. The function $L(v, Z_2)$ pertaining to an atom was obtained as an average over the electron

cloud of the quantity $L(\varrho, v)$ for a gas of density ϱ (Lindhard 1954). As a first approximation, they introduced the expression

$$\begin{aligned} L(v, Z_2) &= \frac{1}{Z_2} \int 4\pi r^2 \varrho(r) \cdot L(r, v) dr \\ &= \frac{1}{Z_2} \int_{r_{min}}^{\infty} 4\pi r^2 \varrho(r) \cdot \log \frac{2mv^2}{\gamma \hbar \omega_o(r)} dr; \frac{2mv^2}{\gamma \hbar \omega_o(r_{min})} = 1 \end{aligned} \quad (6)$$

Here $\varrho(r)$ is the electron density of the target atom, $\omega_o(r)$ is the local plasma frequency $(4\pi e^2 \varrho(r)/m)^{1/2}$; and γ is a constant of the order of $\sqrt{2}$ by means of which they took into account the binding of the electrons. In the Lindhard-Scharff model the mean ionization potential can be calculated from

$$\log I = \frac{1}{Z_2} \int 4\pi r^2 \varrho(r) \log [\gamma \hbar \omega_o(r)] dr \quad (7)$$

Bonderup and Lindhard (1967) and Chu and Powers (1972) calculated I from this formula using Hartree-Fock charge distributions and found significant oscillations superimposed on a slow decrease of I/Z_2 with increasing Z_2 . Even though formula (7) was based on qualitative arguments, the results of the calculations agree fairly well with experiments. It deserves attention that very recently Inokuti et al. (1978) calculated I values by directly using formula (4) in a form appropriate for a continuous distribution of oscillator strengths. They started with the Hartree-Slater central potential model and carried through explicit calculations of df/dE for the entire spectrum from the dipole matrix elements between initial and final electron states. The variation of their I/Z_2 values with Z_2 is similar to that based on the Lindhard-Scharff model which, however, generally gives approximately 30% higher results. This ratio is close to the number $\gamma \sim \sqrt{2}$ which appears as a factor in the I value obtained from formula (7). This result lends strong support to formula (7).

The results based on the Lindhard-Scharff model were so promising even down to velocities of the order of a few times v_o that it was natural to repeat the averaging procedure in formula (6) with a more accurate expression for the electron gas function $L(\varrho, v)$. This was done by Lindhard and Winther (1964), and Bonderup (1967) applied their electron gas results to compute the shell corrections C/Z_2 in formula (3). The first-order Lenz-Jensen distribution was used for the function $\varrho(r)$. The function C/Z_2 is useful in a comparison between theory and experiment since, in contrast to the I value, this quantity is rather insensitive to the details in the distribution of the outermost electrons.

Since the Bethe treatment is based on a first-order perturbation calculation,

the resulting stopping power is proportional to Z_1^2 . However, both range measurements by Barkas (for a survey, see Heckman 1970) and accurate p and α stopping power measurements by Andersen et al. (1969) provided convincing evidence for the existence of higher-order Z_1 correction terms to the stopping formula. Expanding $L(v, Z_2)$ in powers of Z_1 we have the expression

$$L = L_0 + Z_1 \cdot L_1 + Z_1^2 \cdot L_2 \quad (8)$$

where L_0 is given by Eq. (3) whereas L_1 corresponds to a Z_1^3 -correction and L_2 to a Z_1^4 -correction in the Bethe stopping-power formula. Terms of still higher order are neglected.

The Z_1^3 -term, often called the Barkas correction, stems from adiabatic screening effects and receives contributions from both close and distant projectile-electron collisions. An electron gas calculation by Lindhard (1976) and Esbensen (1977) gave the following correction factor

$$Z_1 \cdot L_1 = \frac{3\pi}{2} \cdot \frac{Z_1 e^2 \omega_0}{m \cdot v^3} \cdot L_0 \quad (9)$$

This result is approximately twice that of previous calculations by Ashley et al. (1972, 1973) and by Jackson and McCarthy (1972), both of which neglect the contribution from close collisions.

Bloch's (1933) universal stopping formula which is valid for all values of κ_B describe the transition between Bethe's quantal and Bohr's (1948) classical stopping formula and contains these in the limits of small and large κ_B , respectively. Thus, Bloch's formula contains a correction to the Bethe formula. The correction derives from close collisions only and to first order in κ_B it is given by

$$Z_1^2 L_2 = -1.202 \left(\frac{\kappa_B}{2} \right)^2 = -1.202 \cdot \frac{Z_1^2 v_0^2}{v^2} \quad (10)$$

As pointed out by Lindhard (1976), the Bloch correction is important when comparing p and α measurements.

At low velocities, $v \lesssim v_0 Z_1^{2/3}$, the nuclear as well as the electronic collisions contribute to the slowing-down. The total stopping cross section $S = N^{-1} \left(-\frac{dE}{dR} \right)$ may be written as

$$S = S_e + S_n \quad (11)$$

Simple theoretical considerations lead to velocity-proportional electronic stopping, and a Thomas-Fermi calculation by Lindhard and Scharff (1961) gives

$$S_e = \xi_e 8\pi e^2 a_o \frac{Z_1 Z_2}{(Z_1^{2/3} + Z_2^{2/3})^{3/2}} v/v_o, v < v_o Z_1^{2/3} \quad (12)$$

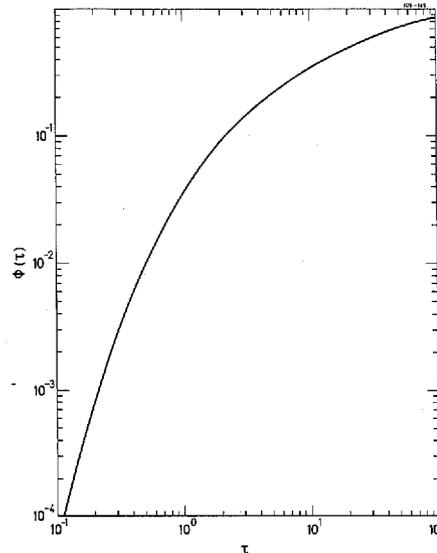
where $\xi_e \sim Z_1^{1/6}$.

In our experimental setup, collisions contributing heavily to the nuclear stopping cause the particles to be scattered out of the angularly narrow forward directed analyzed beam. Hence, only the restricted nuclear-stopping power, denoted by S_n^* , contributes to the measured stopping power. The quantity S_n^* was calculated by Fastrup et al. (1966), and their result may be written as

$$S_n^* = 2.61 \times 10^{-16} \frac{Z_1^2 Z_2^2}{E[\text{keV}]} [M_1/M_2] \Phi(\tau) \left[\frac{\text{eV cm}^2}{\text{atom}} \right]. \quad (13)$$

Here, the dimensionless quantity $\Phi(\tau)$ is a function of $\tau = N\Delta R\pi a^2$ only, where a is the TF screening radius, $a = 0.8853 a_o (Z_1^{2/3} + Z_2^{2/3})^{-1/2}$ and $N\Delta R$ the target thickness. Based on the Lenz-Jensen differential-scattering cross section, the function $\Phi(\tau)$ has been calculated numerically and the result is shown in Fig. 1.

Fig. 1: Function used for calculation of the restricted nuclear stopping power S_n^* . For explanation, cf. text.



§3. Experimental Procedure and Data Treatment

To perform a systematic investigation of energy loss for light ions in gases, we have measured the stopping powers of nine stopping gases (H_2 , He , N_2 , O_2 ,

CO₂, Ne, Ar, Kr, and Xe) for 40-keV to 1-MeV hydrogen and 100-keV to 2.4-MeV helium ions. To cover a large energy region, the measurements were carried out at three different accelerators, an HVEC 2-MV Van de Graaff with magnetic analysis of the energy-degraded beam and a 400-kV Van de Graff and a 100-kV electromagnetic isotope separator both with electrostatic energy analysis. In Table I are shown the energies used at the different accelerators.

Table I

	E_H	E_{He}
2-MV V.d.G.	$0.2 \lesssim E_H \lesssim 1 \text{ MeV}$	$0.2 \lesssim E_{He} \lesssim 2.4 \text{ MeV}$
400-kV V.d.G.	$50 \lesssim E_H \lesssim 300 \text{ keV}$	$100 \lesssim E_{He} \lesssim 300 \text{ keV}$
100-kV sep.	$35 \lesssim E_H \lesssim 70 \text{ keV}$	

The overlap of ion energies investigated with the different machines and analyzing techniques is important since it gives a valuable check on the reproducibility of the experimental data.

The experimental setups are shown in Figs. 2 and 3. After acceleration and momentum analysis in a double-focusing sector magnet, the beam is passed through the differentially pumped target region, energy-analyzed by means of an analyzing magnet or an electrostatic analyzer and detected by a solid-state detector.

A. Stopping cell, gas equipment, and pressure measurement.

The stopping cell is a 504 ± 2 mm long, stainless-steel cylinder of 40-mm diameter. Each end of the gas cell is sealed with brass discs with circular, 0.2-mm diameter apertures. By means of a vacuum feed-through, which allows positional adjustments under vacuum, the gas cell is mounted in a 600-mm long cylinder with 2-mm diameter circular entrance and exit apertures. The pressure in the second differential pumping section and the beam lines is typically around $1\text{-}3 \times 10^{-6}$ torr, while the pressure in the first differential pumping region is $P_1 \simeq 10^{-4} P_G$, P_G being the pressure in the gas cell. The purity of the gases was stated by the commercial supplier (Norsk Hydro) to be better than 99.9%.

It is crucial for obtaining reproducible results that the pressure in the gas cell is kept constant. The stability of the target pressure was maintained via a motor-driven, servo-controlled needle valve (Granville-Phillips Company, Series 216). The gas-cell pressure, $0.1 \lesssim P_G \lesssim 2$ torr, was measured with a membrane manometer (C.G.S. Datametries, type 1083) equipped with a Barocel Pressure Sensor, type 523 H-15. The stated hysteresis was 0.003% and the instrument ranged from zero to two torr with a calibration uncertainty of

0.5%. Owing to the automatic digital readout of the manometer, it was possible to keep a running check of the pressure stability, which was better than 1%.

A mercury thermometer was placed in thermal contact with the vacuum feed-through and thus in contact with the gas cell. The measured temperature was $T = (24.5 \pm 2.5)^\circ\text{C}$ which, to within the stated accuracy, is identical to the target-gas temperature since the localized heating effects caused by the energy dissipation of the passage of the beam through the gas cell can be shown to give rise to a temperature increase of the target gas of less than 0.7°C . This estimate is based on a steady state. In a previous calculation of the localized heating effect by Bourland et al. (1971) the heat conductivity is neglected. Hence a steady state is not established, and their calculation gave too high an increase of the target temperature.

The target thickness $N\Delta R$ (molecules/cm²) is found by integrating the local number density $\mu(x)$ along the beam-path length, i.e.

$$N\Delta R = \int_{-\infty}^{\infty} \mu(x) dx. \quad (14)$$

According to Heinemeier et al. (1975), $\mu(x)$ can be estimated in the following way (i) The pressure is constant within the target cell. (ii) Outside the target cell the density is found as a sum of two terms (a) A constant corresponding to the pressure in the differentially pumped region. (b) A varying term which is equal to the target pressure out to a distance equal to the target aperture and then falls off as the inverse square of the distance. Based on this and assuming the ideal gas law we obtain for the target thickness

$$N\Delta R (\text{mol/cm}^2) = A \frac{273.15}{273.15 + T(^{\circ}\text{C})} \frac{P_G (\text{torr})}{760} \ell_{eff} (\text{cm}) \quad (15)$$

where

$$\ell_{eff} = \ell + 2(r_1 + r_2) + (\ell_1 - \ell) P_1/P_G. \quad (16)$$

$A = 2.6871 \times 10^{19}$ (molecules/cm³) is Loschmidt's number, (ℓ, P_G) and (ℓ_1, P_1) are the length of and the pressure in the gas cell and the first differentially pumped cylinder, respectively, and r_1 and r_2 are the radii of the entrance and exit apertures, respectively. In the present work, the effective length is approximately 0.2% larger than the length of the cell.

B. Energy-analyzing system, detectors, and beam contamination.

The energy-degraded beam was energy-analyzed by a 120-mm radius, 66° cylindrical analyzer (Fig. 2) at the 400-kV Van de Graaff and the 100-kV sepa-

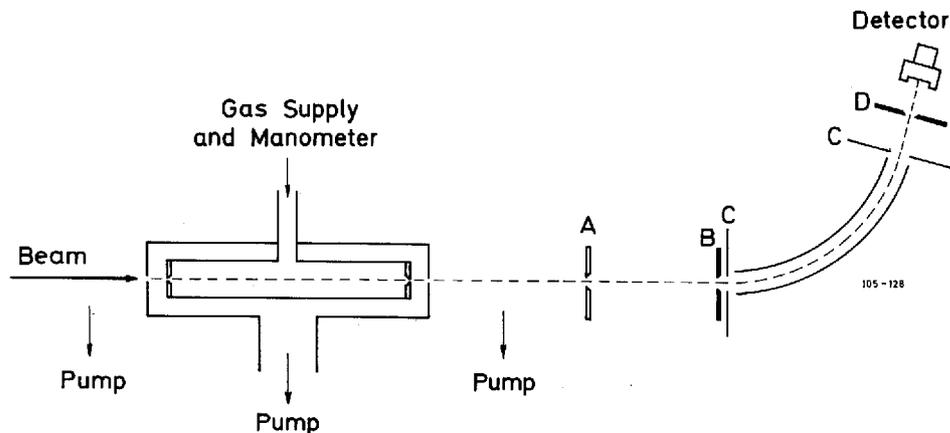


Fig. 2: Experimental setup used at the 400-kV Van de Graaff and at the 100-kV isotope separator.

erator. The analyzer electrodes are 4.9 mm apart and connected to a ~ 12.5 kV symmetric high-voltage supply, which corresponds to the deflection of 0.3 MeV singly charged particles. Apparatures A and B, two 1-mm-wide slits, and aperture D, a 0.42-mm-wide slit, are beam-defining, while the apertures C are used to scrape off scattered halos from the beam.

By extrapolation of the FWHM of the measured energy-loss distributions to zero pressure, we found an energy-independent relative energy resolution (FWHM/E_1) of 0.74%.

The energy calibration was carried out at the 400-kV Van de Graaff accelerator by means of the $^{19}\text{F}(p, \alpha)^{16}\text{O}$ reaction ($E_{res} = 340.46 \pm 0.04$ keV; $\Gamma = 2.4 \pm 0.2$ keV) as the primary standard. The calibration was checked at the 100-kV isotope separator, at which the accelerator voltage was measured directly by a high-voltage probe to within ± 150 V. The measured beam energies and acceleration voltages agreed within the stated accuracy. The analyzer linearity was investigated in connection with measurements of lithium stopping powers (Andersen et al. 1978). Li^+ and Li^{++} beams emerging from krypton and xenon targets much thicker than the mean-free path for charge-exchange processes were analyzed and requiring a difference of a factor of two in the reference voltages, it was found that the measured energies for Li^+ and Li^{++} agreed within $\pm 0.2\%$. The uncertainty in the absolute calibration of the electrostatic analyzer is estimated to be 0.4%.

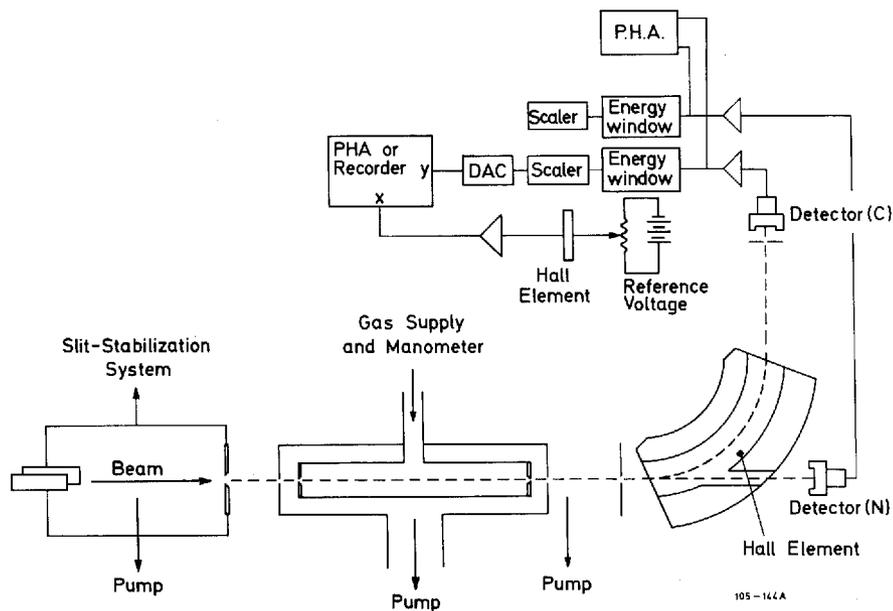


Fig. 3: Experimental setup used at the 2-MV Van de Graaff accelerator.

At the 2-MV Van de Graaff, the energy-loss distributions were analyzed by a 40-cm radius, 90° double-focusing sector magnet (Fig. 3). The magnetic field was measured with a Hall probe, the integral linearity of which was found to be better than 0.5%. The differential linearity of the Hall probe checked by a NMR fluxmeter was better than 0.1%. Tantalum slits with openings of 0.45 mm were placed immediately in front of detector C, and we found a relative energy resolution (FWHM/E_1) of 0.10%. The incoming beam was momentum-analyzed in a 72° double-focusing sector magnet with an energy dispersion of $\Delta E/E \sim 7 \times 10^{-5}$. Hence, in order to obtain a position-stable beam at the target, we used a slit-stabilization system (shown in Fig. 3), which is a feed-back system consisting of two vertical stainless-steel "knives" in front of the target chamber and a set of vertical deflection plates at the exit of the bending magnet.

For the present stopping-power and straggling measurements, the use of solid-state surface-barrier detectors is important. From the detector energy spectrum, it was possible (i) to reveal the presence of low-energy, slit-edge-

scattered particles and (ii) to identify a possible oxygen-beam contaminant with the same kinetic energy as a primary helium beam. The contaminant is formed by electron loss between the base plate and the bending magnet from a $^{16}\text{O}^+$ beam accelerated together with He^+ . Due to the higher pulse-height defect (Steinberg et al. 1972) of the oxygen than that of the helium beam, it is possible to separate the oxygen and helium beams in the pulse-height spectrum. Actually, only once, just after reloading the ion source and under bad vacuum conditions, we did observe any substantial contamination.

C. *Electronic equipment.*

At the 400-keV Van de Graaff and the separator, use of the electrostatic analyzer for the energy analysis enabled us to apply the multiscaling-sweep technique previously employed for lateral spread measurements (Knudsen et al. 1976). A single-channel window was positioned around the main peak in the energy spectrum from the solid-state detector, and the selected signal was fed to the multichannel analyzer, running in multiscaling mode. After adequate biasing and amplification, the horizontal sweep voltage of the multichannel scope was used as an external reference signal for the analyzer high-voltage power supply. The energy window of the analyzer was thus swept over the energy distribution of the beam synchronously with the multiscaler. Through a simultaneous measurement of the reference voltage for the electrostatic analyzer with a digital voltmeter, the energies corresponding to the upper and lower ends of each sweep were determined, and the energy distribution appeared directly as a spectrum in the multichannel analyzer without transformation. Using this sweep technique we did not need any beam normalization, while the experimental equipment used at the 2-MV Van de Graaff and shown in Fig. 3 utilized detector N as a normalizing device.

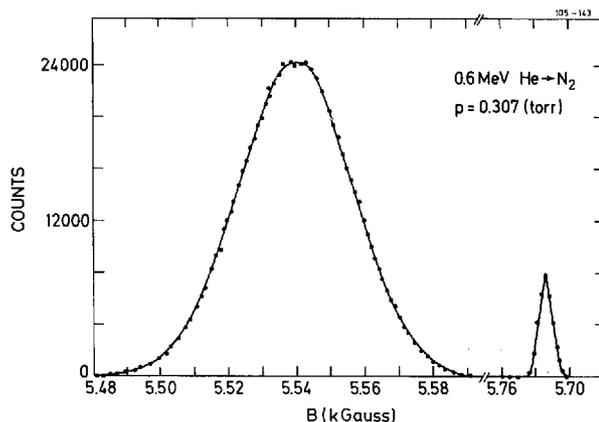
D. *Data treatment.*

With a few exceptions, the pressure in the stopping cell was chosen to make the target thickness satisfy the following criterion,

$$\kappa_V = (\Omega/T_m)^2 = N\Delta R \frac{\pi Z_1^2 Z_2^2 e^4}{4E^2} (M_1/m)^2 > 10. \quad (17)$$

According to Bohr (1948) and Vavilov (1957), Gaussian energy-loss distributions are obtained when inequality (17) is fulfilled. This was confirmed by the measured energy spectra. In the inequality (17), Ω is the standard deviation of the

Fig. 4: Momentum distribution of an incident 600-keV beam and the same beam emerging from an N_2 target.



energy-loss distribution, approximated by the Bohr formula (Bohr 1948), $\Omega_B^2 = 4\pi Z_1^2 Z_2 e^4 N \Delta R$, and T_m is the maximum energy transfer in a single collision with an electron.

With the analyzing magnet, momentum spectra were obtained at each beam energy with and without gas in the stopping cell. A typical momentum distribution is shown in Fig. 4. From the energy-versus-field relation, the energy loss is given by

$$\Delta E = E_i \cdot \frac{\Delta B}{B_i} \left(2 - \frac{\Delta B}{B_i} \right), \quad (18)$$

where $\Delta B/B_i$ is the relative reduction in magnetic field. This shows that for small changes ($\Delta B/B_i \ll 1$), the B axis can be considered as an energy axis. Alternatively, the mean energy loss and the standard deviation are determined from

$$\langle \Delta E \rangle = E_i - (E_1 + E_2)/2 \quad (19)$$

and

$$\Omega_E = \frac{E_2 - E_1}{2\sqrt{2 \log 2}} \quad (20)$$

where E_1 and E_2 are the energies corresponding to the half-maximum positions for the momentum distribution. Formula (19) is preferable to formula (18) in a calculation of $\langle \Delta E \rangle$ due to the larger uncertainty in the determination of the peak position compared to the HWHM positions of the distribution.

With the electrostatic analyzer, determination of the primary-beam energy was more problematic as transmission of the beam without gas in the cell would damage the detector. Hence, the following procedure was adopted: At each

selected beam energy, energy-loss spectra for all gases were measured without changing any of the accelerator settings which may influence the energy, and, at least for one particular gas, energy distributions were measured at three different pressures, corresponding to an energy loss of approximately 5, 9, and 13%. A straight line was fitted to the three measured average energies versus target pressure, and the extrapolation to zero pressure gave a preliminary value of the unattenuated beam energy. Provided the stopping power in question was energy-independent, this value was correct. This not being the case, the preliminary primary energies were used to calculate preliminary stopping powers, and primary energies were obtained through iteration. The uncertainty in the determination of $\langle \Delta E \rangle$ from formula (19) was 1.5% (2σ). Finally, experimental stopping cross sections are obtained as

$$S_o = \frac{\langle \Delta E \rangle}{N \Delta R}. \quad (21)$$

E. Stopping cross sections and corrections.

The electronic-stopping cross section S_e is derived from the observed stopping cross section S_o through subtraction of the restricted nuclear-stopping cross section S_n^* given by formula (13), i.e.,

$$S_e = S_o - S_n^*. \quad (22)$$

For all combinations of target, projectile, and energy, the correction for nuclear collisions is less than 0.5%. However, the correction has been taken into account whenever it exceeds 0.1%.

The energy attributed to the measured stopping cross section is to first order in $\langle \Delta E \rangle / E$ given by

$$E_{av} = E_i - \langle \Delta E \rangle / 2. \quad (23)$$

An expansion by Andersen et al. (1966) of $S(E_{av})$ in powers of $\langle \Delta E \rangle / E$ gives a quadratic correction term to the stopping cross section. As the relative energy loss was always less than 15%, this correction was less than 0.1% and hence omitted.

F. Experimental accuracy.

First, we summarize the quoted systematic errors stemming from uncertainties in the incident-energy E_i (0.3-0.5%), the differential (0.1%), and the integral

(0.5%) linearity of the Hall element, the linearity (0.2%) and the absolute calibration (0.4%) of the electrostatic analyzer, the effective target length (0.4-0.6%) and, finally, the calibration of the membrane manometer (0.5%).

The non-systematic errors originate from uncertainties in the absolute gas temperature ($<1\%$), the pressure in the gas cell ($<1\%$), the HWHM for the degraded energy-loss distribution due to counting statistics (1.5%), and the determination of the length of the sweeping interval (0.3%).

From the uncertainties, all of which correspond to two standard deviations (2σ), it is concluded that the stopping powers are measured to within an uncertainty of 2.5% (2σ).

§4. Experimental Results and Comparison with other Data

The experimental electronic-stopping powers S_e for hydrogen and helium ions in H_2 , He, N_2 , O_2 , Ne, Ar, Kr, Xe, and CO_2 are plotted in Figs. 5-18 as functions of E_{av} . In the figures, the present results have been compared with most other published hydrogen and helium energy-loss data. The scatter of our data points is consistent with the estimate of the measuring accuracy given above.

A. Hydrogen stopping powers.

Recently Andersen and Ziegler (1977) published tabulations of hydrogen stopping powers for all elements in the energy range $10 \text{ keV} < (E/\text{amu}) < 20 \text{ MeV}$. These semi-empirical stopping-power fits are plotted in Figs. 5-13. From the figures we first note that for most of the targets used, good agreement exists between the Andersen and Ziegler semi-empirical stopping-power fits and the present experimental results for energies $E \gtrsim 100 \text{ keV}$. However, the semi-empirical fits have a tendency of being slightly low around the stopping-power maximum, and for Xe targets, the fit to our data as well as to those previously published is rather poor over a broad energy range. The present results agree within the stated accuracy with the averaged S_H -values found by Reynolds et al. (1953), the accuracy of which is 2-4% (2σ), while the data obtained by Phillips (1953) are systematically $\sim 15\%$ lower. Since Phillip's results are included in the data on which the tabulations by Andersen and Ziegler are based, their curve appears to be too low for energies $E \lesssim 100 \text{ keV}$.

Fig. 5: Stopping-power results for hydrogen in H_2 .

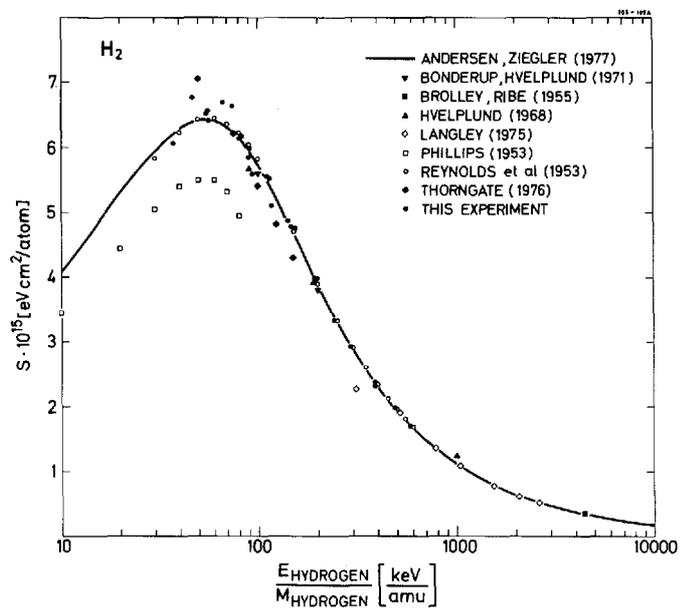


Fig. 6: Stopping-power results for hydrogen in helium.

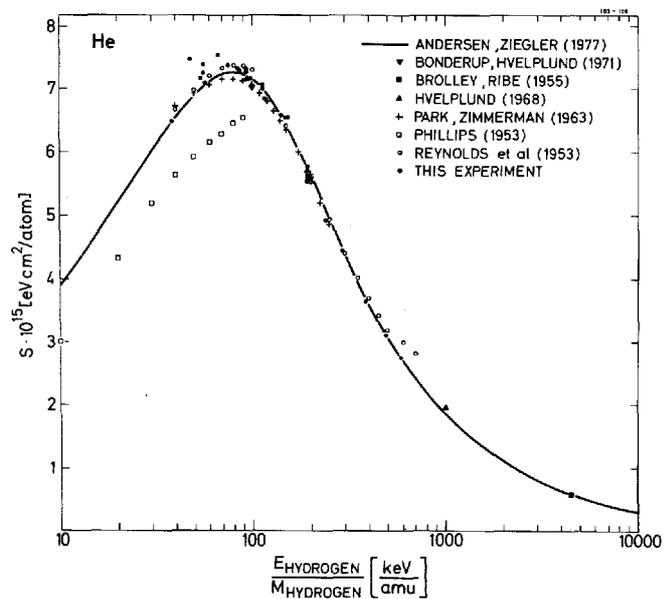


Fig. 7: Stopping-power results for hydrogen in N_2 .

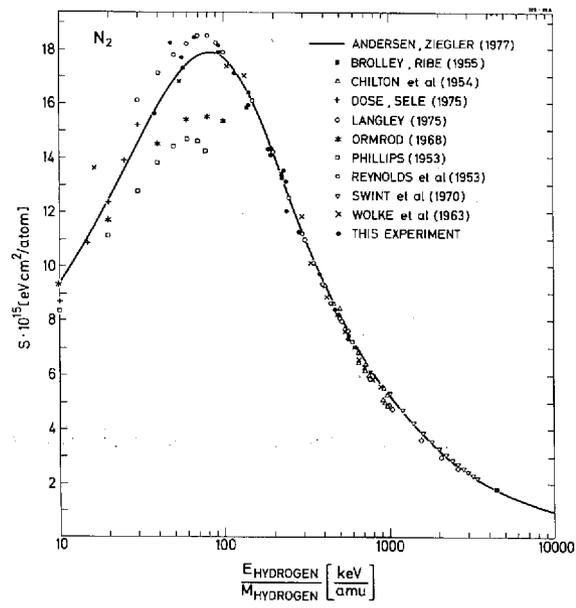


Fig. 8: Stopping-power results for hydrogen in O_2 .

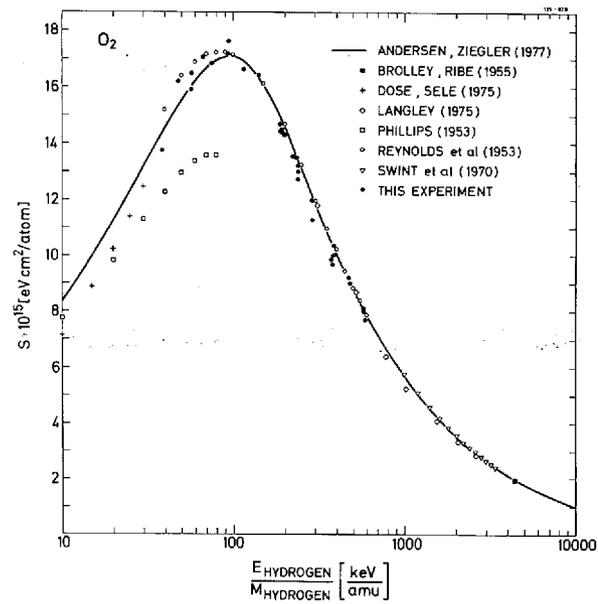


Fig. 9: Stopping-power results for hydrogen in neon.

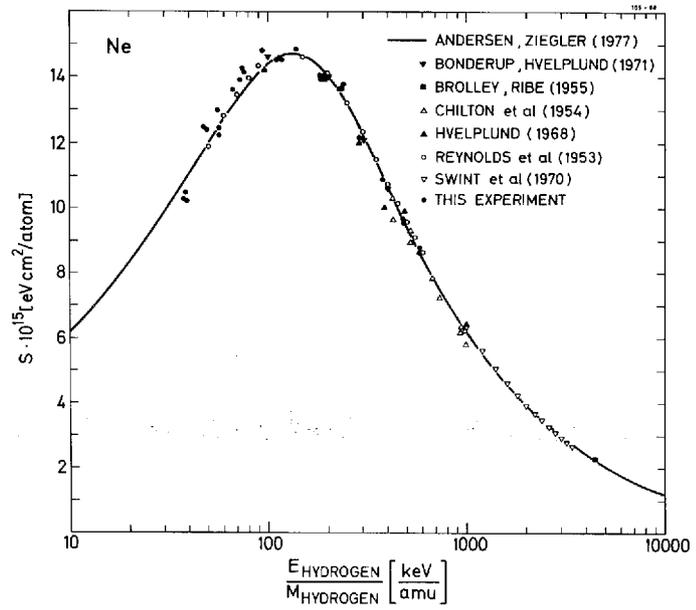


Fig. 10: Stopping-power results for hydrogen in argon.

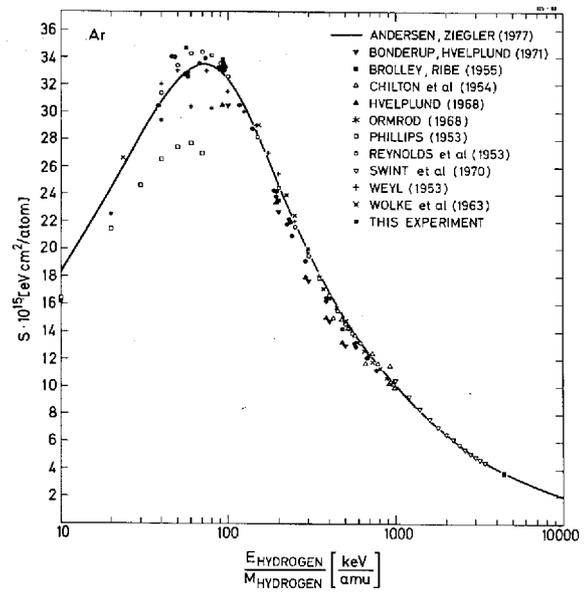


Fig. 11: Stopping-power results for hydrogen in krypton.

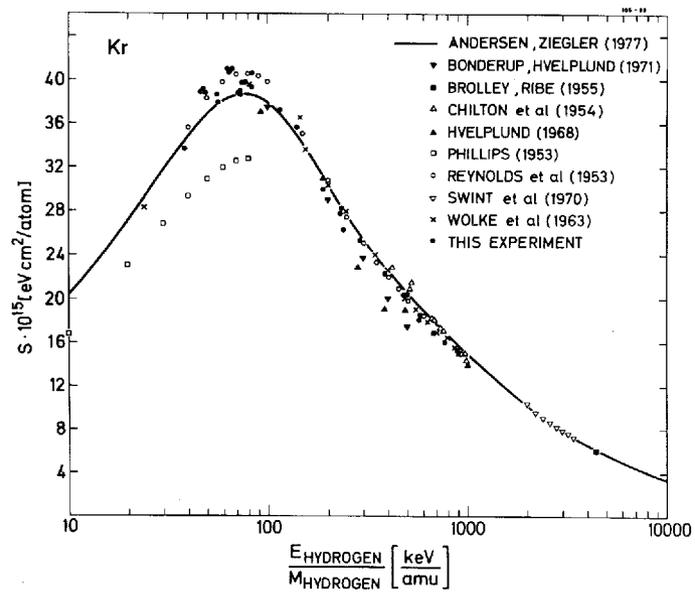


Fig. 12: Stopping-power results for hydrogen in xenon.

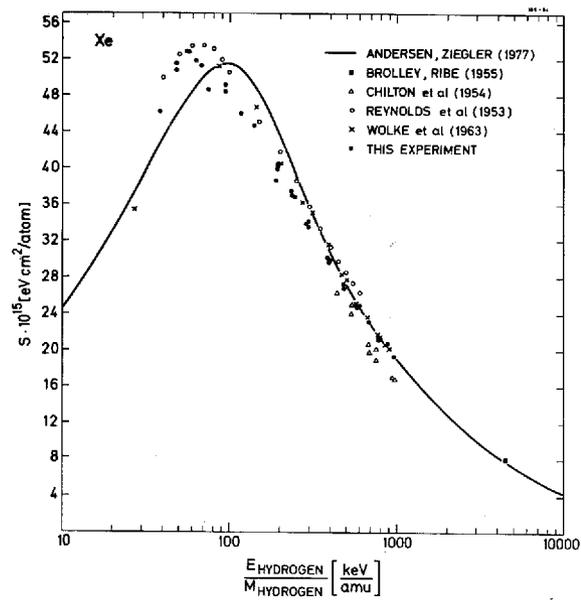
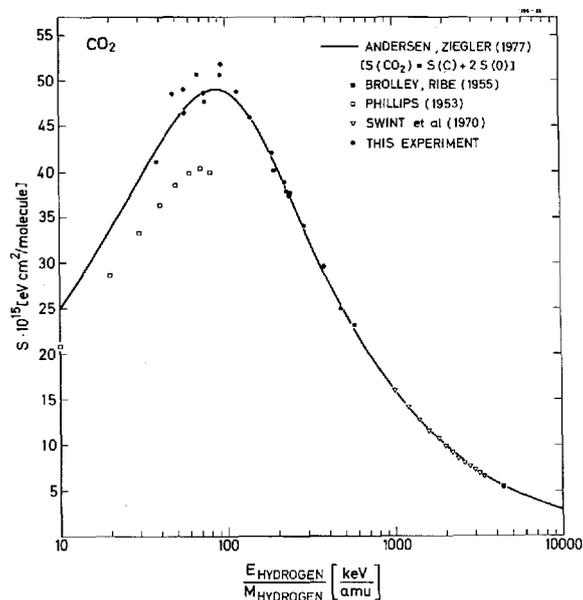


Fig. 13: Stopping-power results for hydrogen in CO_2 .



B. Helium stopping powers.

Very recently, Ziegler (1978) published helium stopping-power tabulations similar to those of Andersen and Ziegler (1977) mentioned above. As the present experimental results are included in Ziegler's helium-data base, they constitute no independent check of these tables. Hence the tabulations are not included in Figs. 14-18. Nor are the polynomial fits made by Ziegler and Chu (1974) since for the present gases, these fits are identical to the averaged S_{He} values of the Baylor group (see below) as plotted in Figs. 14-18.

With a setup more or less equivalent to the present one, the Baylor group, Bourland, Chu, and Powers (1971) and Chu and Powers (1971) have made a systematic investigation of α -particle stopping cross sections in gases. The stated accuracy of their measurements ranges from 1.5% to 3% (2σ). As can be seen from the figures, the measurements performed by the Baylor group are higher by 1-6% than the present S_{He} results. This difference is not understood at present. However, it should be pointed out that the Baylor group employed a non-energy dispersive detector (a Faraday cup), and hence could not reveal the presence of slit-edge-scattered particles and/or a possible oxygen contamination of the helium beam. As they have not published the straggling results, it is not possible to disclose whether beam contamination and/or pressure fluctuations were significant. Finally, as a McLeod gauge was used, the pressure could not

be checked continuously. However, it should be noted that the Baylor group performed independent measurements in a sealed gas cell, using a solid-state detector. These measurements ($\pm 5.4\%$ to $\pm 6.7\%$ (2σ)) were not as accurate as the differentially pumped gas-cell measurements but they did agree with the latter to within the stated accuracy.

In this connection, it should be noted that the reliability of the present measurements is enhanced due to the fact that the stopping cross sections for both hydrogen and helium ions were measured in the same gases, and the present S_H data agree with most other S_H data.

The stopping cross sections by Hvelplund (1968, 1971) for helium in H_2 , He, O_2 , and Ne are systematically 10-15% below the present data. This discrepancy is difficult to understand since Hvelplund's equipment was similar to that used here. However, it is fair to mention that the data by Fastrup et al. (1968), and the lithium stopping power results by Andersen et al. (1978) lie 10-16% and 20% above Hvelplund's results, respectively. Hvelplund's proton stopping-power and straggling data (Bonderup and Hvelplund 1971) are in much better agreement with our present data.

Using a natural α -emitter source and a solid-state detector, Hanke and Bichsel (1970), Palmer (1966), and Rotondi (1968) measured range-energy relations for α particles in Ar, H_2 , CO_2 , N_2 , and O_2 . Hanke and Bichsel made a thorough data evaluation, taking into account corrections for multiple scattering, discrete energy loss (Lewis correction), undetected energy losses in the detector gold-surface layer, and adjoining dead layer and energy loss due to self-absorption in the source head. Thus, by differentiating the range-energy curve, they determined the stopping cross section for helium in argon with a claimed accuracy of 0.2% for $E_{He} > 2$ MeV, decreasing to 1.5% for $E_{He} \sim 1$ MeV. Palmer and Rotondi did not correct their range-energy results, and this may explain why their stopping-power results, especially around the maximum, deviate significantly from the present data.

By using an ^{241}Am α source and a solid-state detector, Kerr et al. (1966) and Wenger et al. (1973) have measured S_{He} at energies $0.3 \lesssim E_{He} \lesssim 5$ MeV by either varying the pressure in the absorber chamber between 0 and 720 mm Hg (Kerr et al. 1966) or by changing the distance between source and detector with the gas held between them kept at a pressure of 2.54 torr above atmospheric pressure (Wenger et al. 1973). Since there is a large uncertainty in the energy, which they attribute to the measured stopping power due to the large energy loss and as the above corrections to the stopping power are not taken into account, it is evident that these measurements are vitiated by large uncertainties. This is also the case for the data by Ramirez et al. (1969) for similar reasons.

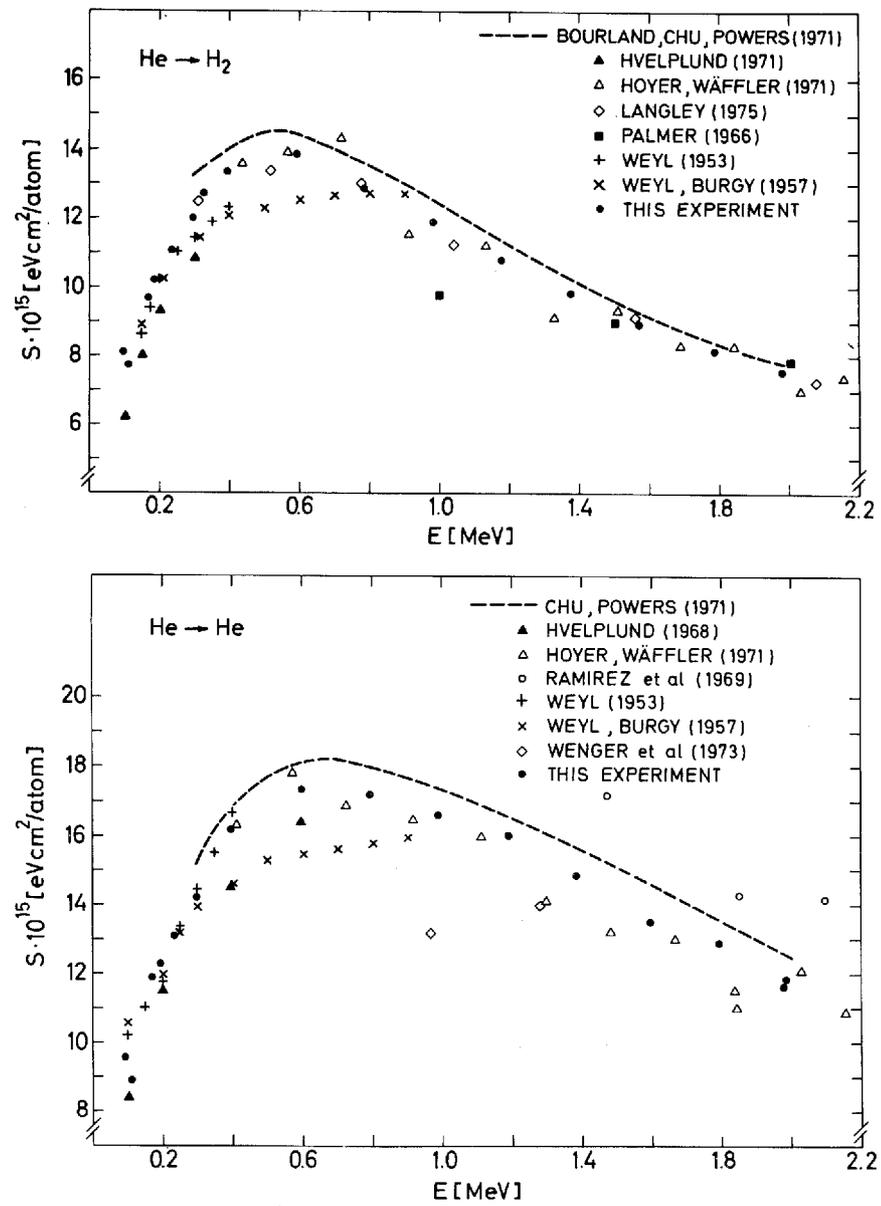


Fig. 14: Stopping-power results for helium in H₂ and He.

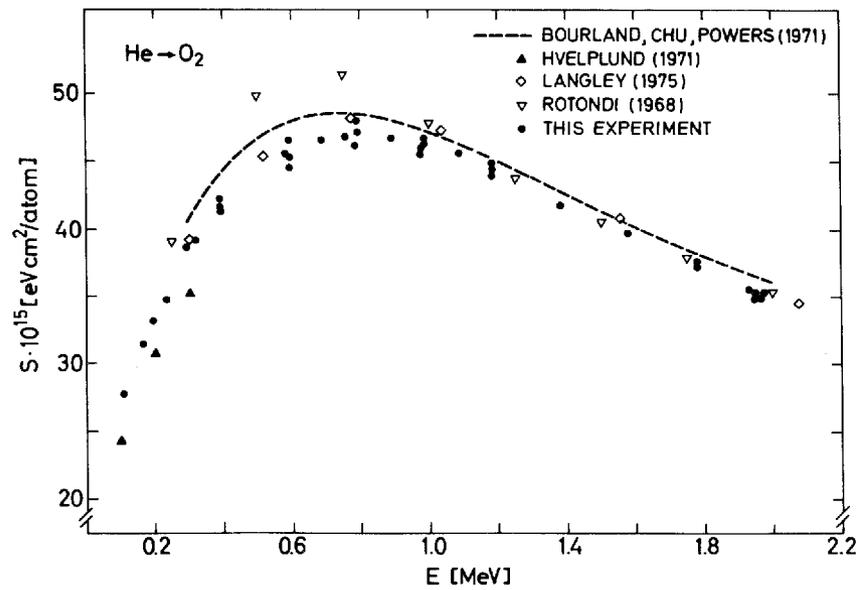
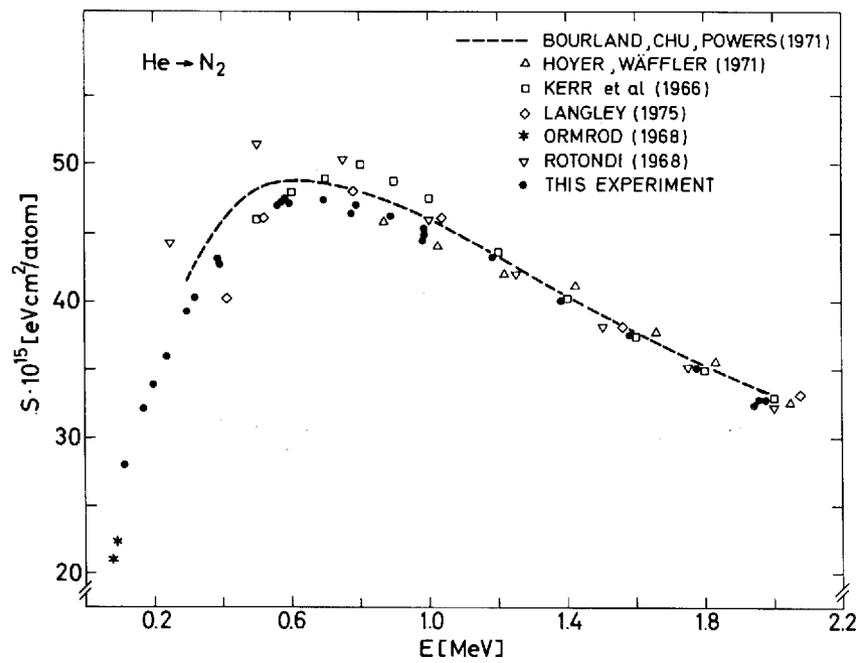


Fig. 15: Stopping-power results for helium in N₂ and O₂.

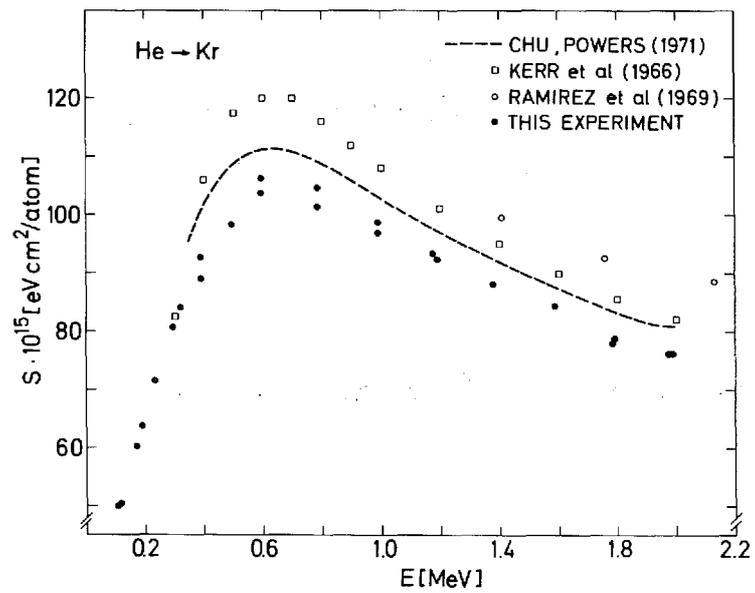
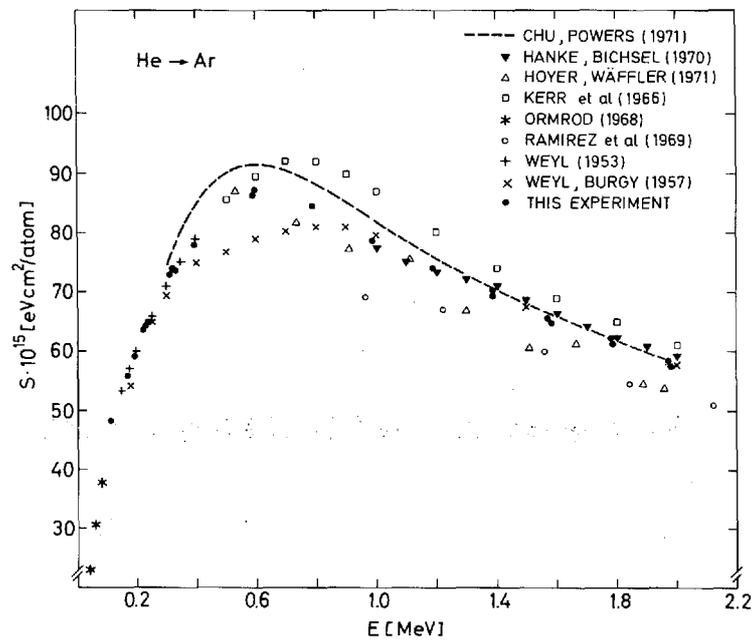


Fig. 16: Stopping-power results for helium in argon and krypton.

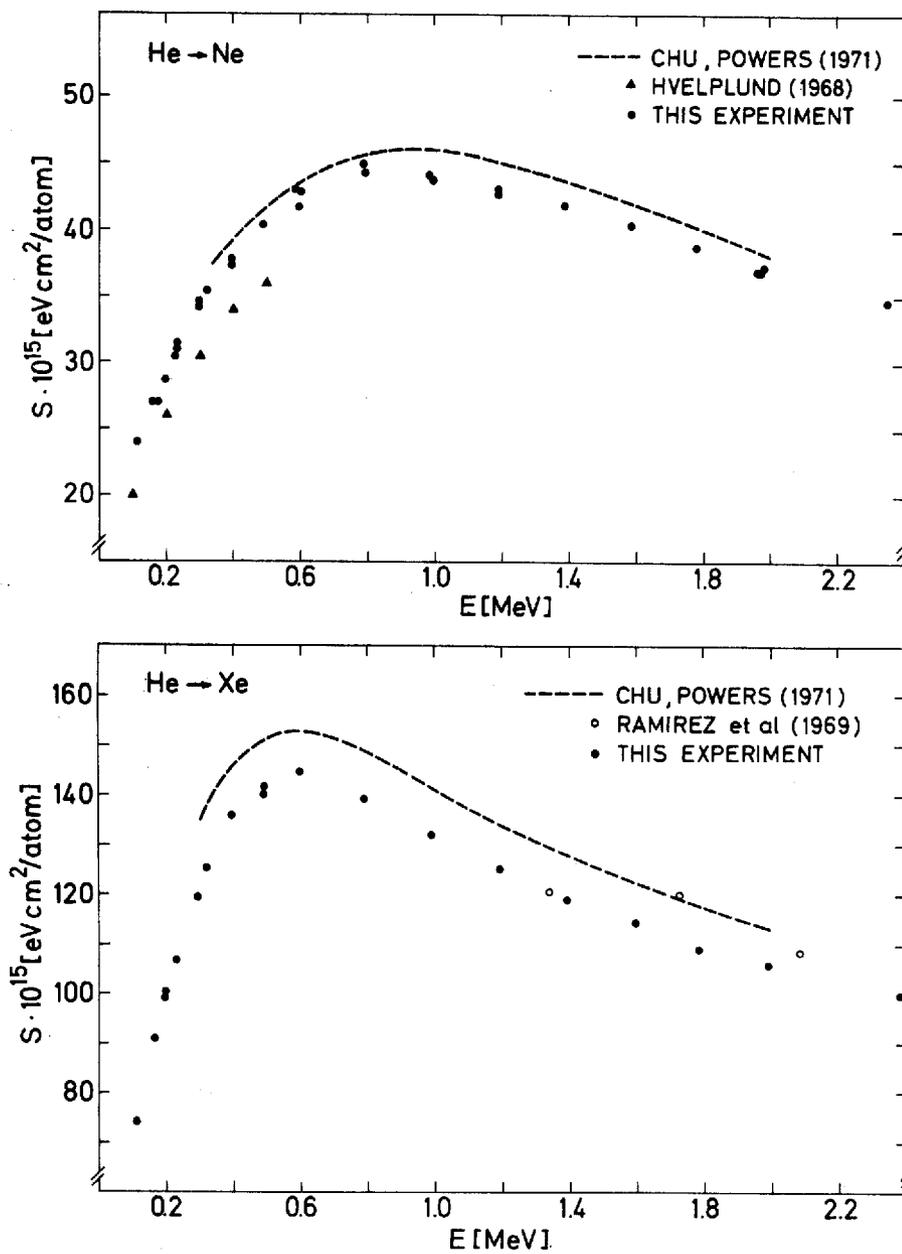


Fig. 17: Stopping-power results for helium in neon and xenon.

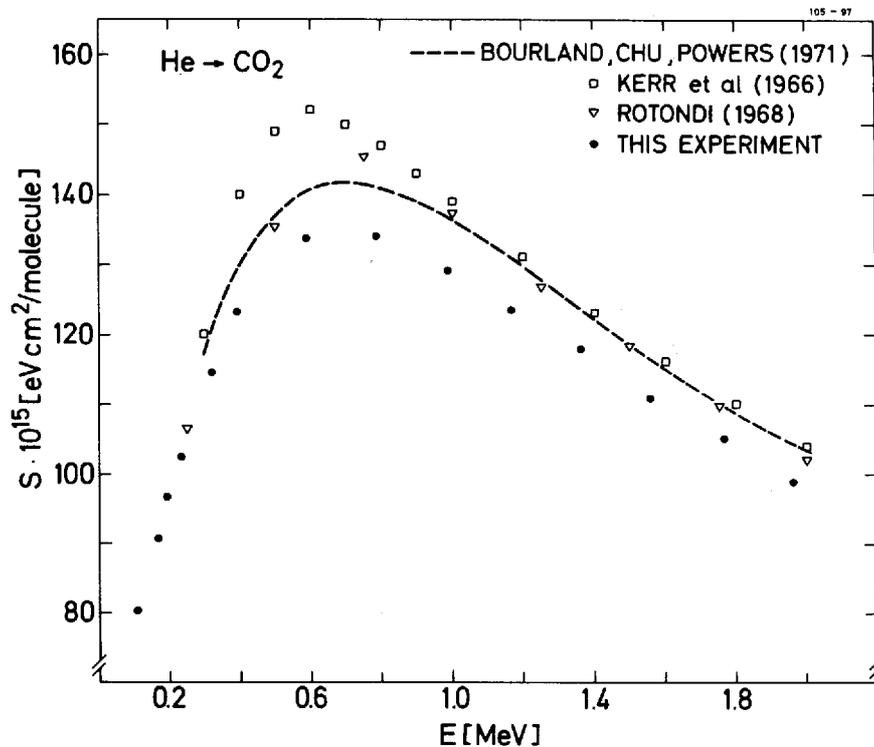


Fig. 18: Stopping-power results for helium in CO₂.

§5. Comparison with Theory

A. Shell corrections and I values.

Before making any comparison of the experimental results with the calculations based on the Lindhard-Scharff model and its refinements, we should emphasize that the region of validity of a perturbation treatment for a free electron gas is restricted to $\kappa_B < 1$ or $v > 2v_0 Z_1$ (Bonderup 1967), i.e., to hydrogen and helium energies higher than ~ 100 keV and ~ 1600 keV, respectively. Therefore any comparison between the present helium stopping powers and existing calculations should be taken with some reservation.

As argued in §2, it is advantageous to apply the so-called shell corrections C/Z_2 rather than $L(v, Z_2)$ in a detailed comparison between theory and experiment. Introducing Bichsel's (1964) reduced stopping-power variable,

$$X = \log \left(\frac{2mv^2}{1 - (v/c)^2} \right) - \left(\frac{v}{c} \right)^2 - 1/\zeta_2 \frac{mv^2}{4\pi e^4 \zeta_1^2} S_{\text{exp}}, \quad (24)$$

we may deduce empirical shell corrections from the experimental hydrogen stopping powers using formulas (2) and (3) as

$$(C/\zeta_2)_{\text{exp}}^* = X_{\text{exp}}^H - \log I. \quad (25)$$

As will be discussed later the shell corrections deduced from this formula are not genuine shell corrections and are therefore labelled with an asterix.

In Figs. 19 and 20, experimental shell corrections $(C/\zeta_2)^*$ obtained from formula (25) are presented as a function of energy. The I values used are those extracted by Andersen and Ziegler (1977) from their fits to previously published data. Comparison is made to Bonderup's (1967) theoretical calculations $(C/\zeta_2)_B$ and to the empirical shell corrections by Andersen and Ziegler. As an indication of the sensitivity of $(C/\zeta_2)^*$ to uncertainties in the S_H data, the distance between the dashed line "3% effect on stopping" and the zero line gives the change in $(C/\zeta_2)^*$ due to a 3% change in stopping power. An increase in stopping power gives a lower X value and hence a lower $(C/\zeta_2)^*$ value.

From Figs. 19 and 20 it is observed that the experimental $(C/\zeta_2)^*$ data deviate significantly from Bonderup's shell corrections $(C/\zeta_2)_B$ both in size and shape (see, e.g., $H \rightarrow Ne$). Until recently, it was believed that this discrepancy was due to a deficiency in the theoretical model in the energy region discussed here. However, all possible deviations from formulas (2) and (3) such as higher-order ζ_1 corrections are automatically included in the empirical $(C/\zeta_2)^*$ values. According to formula (8), the correct shell correction is determined from

$$(C/\zeta_2) = X_{\text{exp}}^H - \log I + \zeta_1 L_1 + \zeta_1^2 L_2. \quad (26)$$

Therefore the experimental $(C/\zeta_2)^*$ proton values in Figs. 19 and 20 should be compared to the following "apparent" theoretical shell-corrections

$$(C/\zeta_2)_{th}^* = (C/\zeta_2)_B - (L_1 + L_2). \quad (27)$$

In calculating $(C/\zeta_2)_{th}^*$, we have used the Bloch expression, formula (10), for L_2 . According to the discussion in connection with formula (9), the Barkas correction L_1 has been set equal to twice the quantity given by Jackson and McCarthy (1972)

$$L_1 = 2L_1^{JM} = 2 \frac{F\left(\frac{v}{v_0} \frac{1}{\zeta_2^{1/2}}\right)}{\zeta_2^{1/2}} L_0. \quad (28)$$

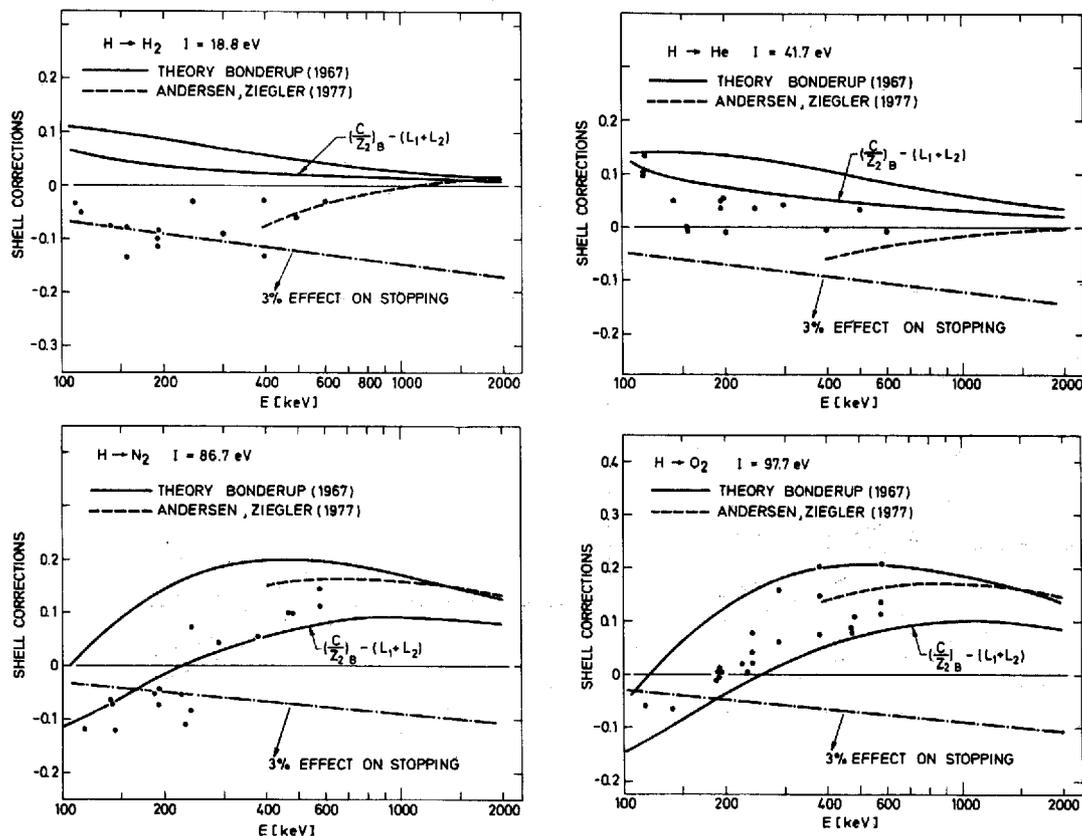


Fig. 19: Experimental shell corrections for H_2 , He, N_2 , and O_2 obtained from Eq. (25) (points), compared with Andersen and Ziegler's semiempirical fit (dashed), Bonderup's theoretical values and those values corrected for higher-order Z_1 effects, according to Eq. (27). The I values are those given by Andersen and Ziegler (1977).

In calculating L_1 for the heavier elements, we have extrapolated L_1^{JM} slightly outside the stated region of validity given by $\frac{1}{Z_2^{1/2}} \frac{v}{v_0} \gtrsim 0.8$. Theoretical shell corrections obtained from formula (27) are also plotted in Figs. 19 and 20.

In the cases of N_2 , Ne, Ar, and Xe, the corrected values $(C/Z_2)_{th}^* = (C/Z_2)_B - (L_1 + L_2)$ show perfect agreement with the experimental values $(C/Z_2)^*$. For H_2 , He, O_2 , and Kr, the energy dependence of the experimental data $(C/Z_2)^*$ agrees much better with the energy dependence of the corrected values

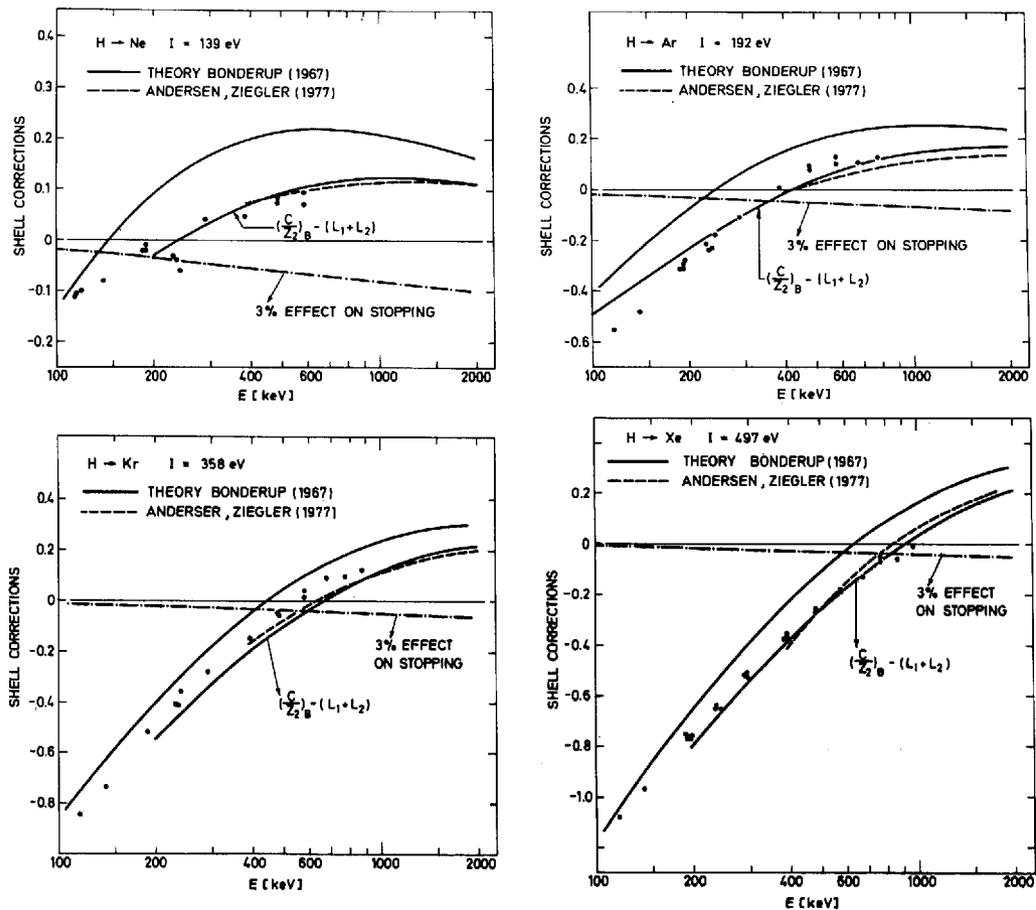


Fig. 20: Experimental shell corrections for Ne, Ar, Kr, and Xe. See caption to Fig. 19 for further explanation.

$(C/Z_2)_{ih}^* = (C/Z_2)_B - (L_1 + L_2)$ than with that of $(C/Z_2)_B$. However, there is a systematic shift in the absolute value between $(C/Z_2)^*$ and $(C/Z_2)_{ih}^*$, which may originate in the choice of I value. A change of I to $I - \Delta I$ changes the experimental $(C/Z_2)^*$ data by the additive amount $\Delta I/I$. In this way, experimental I values are determined from the present data. In Table II, the resulting experimental ionization potentials are compared with those given by Andersen and Ziegler (1977) and Chu and Powers (1972).

Table II

$I(\text{eV})$	H ₂	He	N ₂	O ₂	Ne	Ar	Kr	Xe
Present results	17.6	40.7	86.7	102.1	139	194	376	497
Andersen-Ziegler	18.8	41.7	86.7	97.7	139	194	358	497
Chu-Powers			92.4	110	160	207	403	529

It is concluded that the present I values, which constitute an independent check of the Andersen and Ziegler I values, are in good agreement with these, while the theoretical I values by Chu and Powers are systematically too high.

Above, we used the experimental shell corrections $(C/\zeta_2)^*$ as a standard and corrected Bonderup's theoretical calculations $(C/\zeta_2)_B$ for comparison. Andersen et al. (1977) employed $(C/\zeta_2)_B$ as a standard and corrected $(C/\zeta_2)^*$ by adding $(L_1 + L_2)$ for comparison. The reason for our approach is that in the present energy region, we are not able to determine L_1 and L_2 experimentally as was done by Andersen and coworkers and hence we could not obtain experimentally determined genuine shell corrections with which to compare the theoretical $(C/\zeta_2)_B$ results.

A different approach is to use the empirical fits for L_1 and L_2 extracted from the measurements by Andersen et al. (1977) and denoted by L_1^A and L_2^A , respectively. For $\zeta_2 > 10$, good agreement is found between $(C/\zeta_2)_B - (L_1 + L_2)^A$ and formula (27) shown in Figs. 19 and 20, while the agreement becomes increasingly unsatisfactory for $\zeta_2 < 10$. As Andersen and coworkers measured stopping powers of Al, Cu, Ag, and Au, their empirical fits for L_1 and L_2 should be employed only for $13 \lesssim \zeta_2 \lesssim 79$, in which case they give results which are consistent with the present data.

Concerning Andersen and Ziegler's (1977) "fitted shell corrections", $(C/\zeta_2)^{AZ}$, all higher-order ζ_1 contributions are piled onto these, and $(C/\zeta_2)^{AZ}$ should therefore be compared with the present results $(C/\zeta_2)^*$ or with $(C/\zeta_2)_B - (L_1 + L_2)$. In the cases of Xe, Kr, Ar, Ne, and O₂, there is good agreement between $(C/\zeta_2)^{AZ}$ and the present data $(C/\zeta_2)^*$. For N₂, the curve for $(C/\zeta_2)^{AZ}$ is somewhat higher than $(C/\zeta_2)^*$, while for H₂ and He, the fit $(C/\zeta_2)^{AZ}$ deviates both in magnitude and energy dependence from the present results. If Andersen and Ziegler had based their fit for H₂ and He not on the scarce and scattered available experimental data points but had extrapolated the shell-correction fit obtained for other elements with $\zeta_2 \geq 3$ to the cases of H₂ and He, the present H₂ and He results would have agreed with the general trend on the $(C/\zeta_2)^{AZ}$ curves.

B. Experimental determination of the Z_1^3 correction.

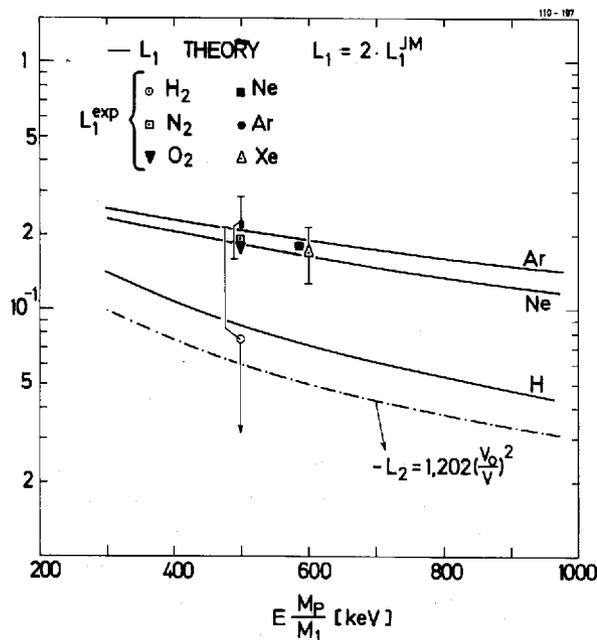
We assumed above that the Barkas and the Bloch corrections were given by the formulas (28) and (10), respectively. Andersen et al. (1977) have demonstrated that this is approximately correct for $7 \lesssim v/v_0 \lesssim 12$. In our case we can extract information on the higher order Z_1 terms only when the helium ions are known not to carry electrons in a bound state. Equilibrium charge-state measurements show that this is the case for $E_{He} \gtrsim 1.6$ -2 MeV. Furthermore, from measurements with hydrogen and helium ions only, it is not possible to distinguish between Z_1^3 and Z_1^4 corrections. However, based on the results by Andersen et al. (1977), it seems reasonable to assume that L_2 is given approximately by the Bloch expression, formula (10). As follows from Eqs. (8) and (24), experimental L_1 values may then be extracted from the formula

$$L_1^{exp} = X_H^{exp} - X_{He}^{exp} - 3L_2. \quad (29)$$

The uncertainties in L_1^{exp} are calculated under the assumption that the uncertainty in L_2 is $\sim 20\%$.

In Fig. 21, experimental L_1 values for H_2 , N_2 , O_2 , Ne, Ar, and Xe are compared with the theoretical estimate for L_1 given by Eq. (28), which is a good

Fig. 21: Experimental results for the Z_1^3 factor L_1 obtained from Eq. (29). Theoretical curves for L_1 (solid) and L_2 (dot-and-dash) are Eqs. (28) and (10) respectively.



approximation to the results of Lindhard (1976) and Esbensen (1977). From Fig. 21, it is concluded that the experimental L_1 values are in good agreement with the expression $L_1 = 2L_1^{JM}$. This supports Lindhard's and Esbensen's value for the Z_1^3 correction and thereby the L_1 expression used in the preceding section in connection with the discussion of shell corrections. It should be noted that the deviation of L_1^{exp} from L_1^{JM} cannot be explained as a charge-state effect since the introduction of an effective charge state $Z_{He}^* < 2$ would increase the L_1^{exp} values in Fig. 21.

Ranges of Σ^+ and Σ^- in emulsion for $v \sim 20v_0$ were found to differ by an amount corresponding to the value given by Lindhard (1976), while the range differences in hydrogen were a factor of five larger than Lindhard's prediction (for a review of the experimental range data, see Heckman (1970)). On this background it is important that the present experimental L_1 for hydrogen agrees with Lindhard's findings. The large uncertainty in $L_1(\text{H}_2)$ is due to the high $L(v, Z_2)$ value.

Only Ward et al. (1976) have previously analyzed their data for Z_1^3 effects for $v/v_0 < 5$. They measured stopping cross sections for hydrogen and helium in aluminum and found that, within the experimental uncertainty, the quantity $(S_{He} - 4S_H)/S_{He}$ was equal to zero for $v/v_0 \sim 4.5$. From this they concluded that no Z_1^3 effect was present. However, taking into account the Z_1^4 correction, we find

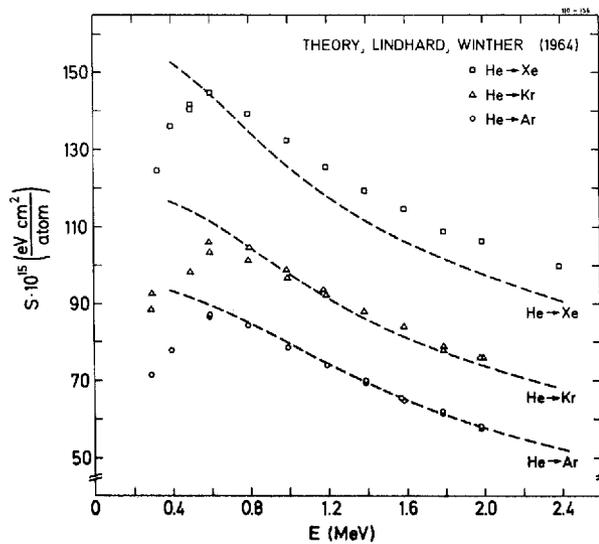
$$\frac{S_{He} - 4S_H}{S_{He}} = \frac{L_1 + 3L_2}{L + 2L_1 + 4L_2} = 0.84\% \text{ for } v/v_0 = 4.5.$$

This is in perfect agreement with the results in Fig. 12 of the paper by Ward et al., and their data hence confirm the magnitude of the higher-order Z_1 effects found in this work.

C. Helium stopping powers.

As mentioned above, it is not possible to make any detailed comparison between S_{He} data and perturbation calculations for a free electron gas at the present energies. In spite of this, Rousseau et al. (1971) have used the Lindhard-Winther expressions for an electron gas with the charge-densities obtained from Hartree-Fock-Slater wave functions to calculate the stopping cross section for 0.4-2 MeV α particles. The calculations, with which Chu and Powers (1971) compare their noble-gas results, are wrong due to problems with the joining of the asymptotic expressions by Lindhard and Winther. Comparing the present S_{He} results with later and corrected calculations by Ziegler and Chu (1974), we find surpris-

Fig. 22: Stopping cross sections for He penetrating Ar, Kr, and Xe compared with calculations by Ziegler and Chu (1974) based on Lindhard-Winther theory.



ingly good agreement (to within $\pm 7\%$ even at energies as low as $E_{He} \sim 600 \text{ keV}$). Figure 22 shows the situation for Xe, Kr, and Ar.

D. The ratio S_{He}/S_H as a function of velocity.

As mentioned above, a quantal perturbation calculation of the stopping cross section is restricted to velocities $v > 2v_0 Z_1$. At lower velocities, it has, nevertheless, been suggested that the usual stopping formula (2) be applied with the charge number Z_1 replaced by an effective charge number, i. e.

$$S = \frac{4\pi e^4 Z_2}{mv^2} (Z_1^*)^2 L(v, Z_2). \quad (30)$$

Many authors have used this approach to analyze experimental stopping-power data in terms of charge states, comparing heavy-ion stopping powers with corresponding proton-stopping powers at the same velocity. It should, however, be emphasized that no theoretical basis exists for this charge-state scaling procedure.

In Fig. 23, the ratio between the present hydrogen and helium stopping powers measured at the same velocity are shown as smooth curves. These data lead to the following conclusions:

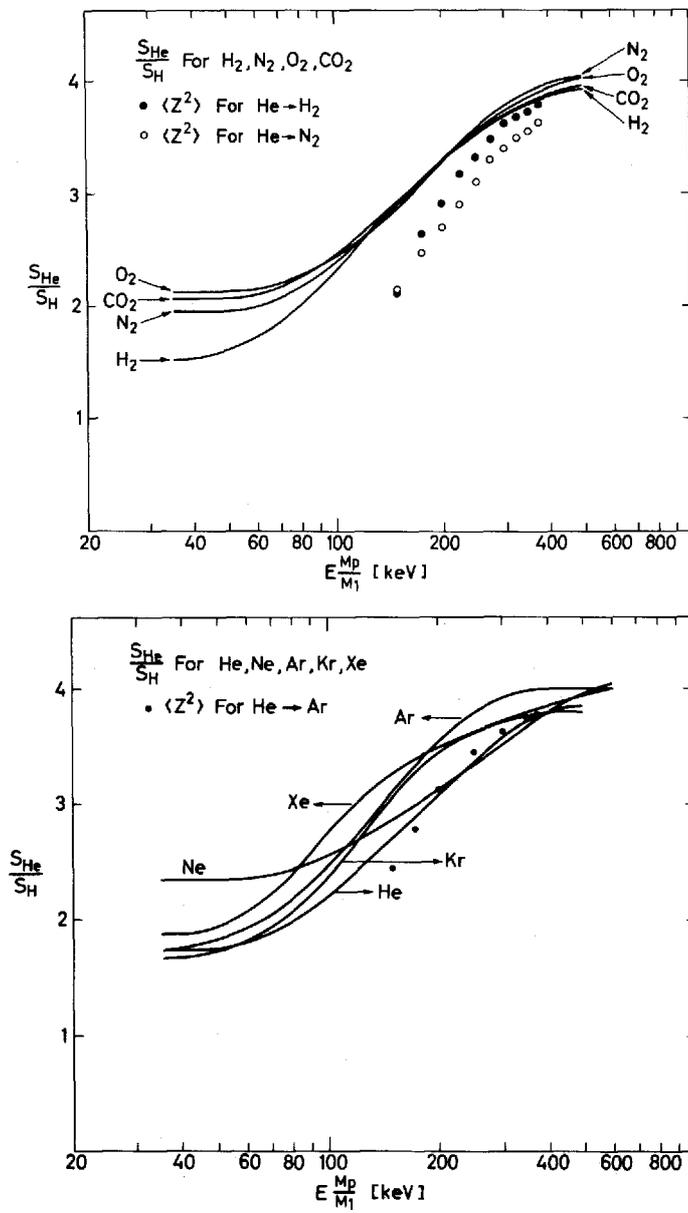


Fig. 23: Experimental stopping-power ratios (solid curves) of He ions to hydrogen ions evaluated at the same velocity, compared with Pivovar's (1961) measured mean-square charge state for He ions in H_2 , N_2 , and Ar.

(1) The ratios S_{He}/S_H are not independent of Z_2 , as was also found by Sautter and Zimmermann (1965). Thus it is not possible, at least not for gases, in a simple way to scale $S_{He}(v, Z_2)$ to $S_H(v, Z_2)$ or vice versa.

(2) From being a constant nearly equal to four for $E/\text{amu} > 500$ keV, S_{He}/S_H decreases with decreasing velocity and approaches another constant value for $E/\text{amu} \lesssim 50$ keV, and this value depends strongly on Z_2 . If the stopping power in the low energy region is written in power form as $S = kE^p$, k and p being constants, the exponents p are approximately equal for hydrogen and helium ions in a given target material. According to the wellknown Lindhard-Scharff (1961) velocity-proportional stopping formula (12), the constant ratio is given by

$$(S_{He}/S_H)^{LS} = 2^{7/6} \frac{(1 + Z_2^{2/3})^{3/2}}{(2^{2/3} + Z_2^{2/3})^{3/2}}. \quad (31)$$

In Table III, the experimental ratios are compared with the L - S values for $E/\text{amu} \lesssim 50$ keV. Rather good agreement is found for the lighter target elements.

Table III

	H ₂	He	N ₂	O ₂	Ne	Ar	Kr	Xe
exp.	1.53	1.75	1.95	2.15	2.35	1.75	1.65	1.88
L - S	1.53	1.65	1.88	1.90	1.93	2.02	2.09	2.12

It might be noted that our lithium stopping-power results for the same gases (Andersen et al. 1978) at energies $25 \text{ keV} \lesssim E/\text{amu} \lesssim 75 \text{ keV}$, give exponents p which are systematically higher than 0.5 while those for helium are lower than 0.5 (see Table IV in the next section). Thus the S_{Li}/S_{He} ratios are energy-dependent.

(3) From formula (30), the ratio S_{He}/S_H is given by

$$S_{He}/S_H = (Z_{He}^*)^2 / (Z_H^*)^2 \times L_{He}/L_H. \quad (32)$$

Since $L_{He}/L_H \simeq 1$ and $(Z_H^*)^2 \simeq 1$ for $E_H \gtrsim 150$ keV, one would expect that $S_{He}/S_H \simeq (Z_{He}^*)^2$ down to $E(M_p/M) \gtrsim 150$ keV. In Fig. 23, the (S_{He}/S_H) ratios in H₂, N₂, and Ar are compared with the mean-square charge states

$$\langle Z^2 \rangle = \sum_i i^2 \times F_{i\infty} \quad (33)$$

for helium in H₂, N₂, and Ar obtained from equilibrium charge-state measurements by Pivovar et al. (1961). $F_{i\infty}$ denotes the equilibrium charge-state fraction

of the beam in charge state i . A discrepancy of the order of 10% is revealed for $E/\text{amu} \lesssim 375$ keV.

It is concluded that at the one percent level the helium stopping powers, at least for gases, are inconsistent with Eq. (30) with $(Z_1^*)^2$ equal to the mean-square charge state measured directly, and the present results support the theoretical reservation against Eq. (30). The parameter $(Z_1^*)^2$ is no more than a scaling parameter which, on the other hand, appears to be useful for a prediction of unmeasured heavy-ion stopping powers with a required accuracy in the 15 percent range.

Concerning the ratios S_{He}/S_H , it should be emphasized that the S_{He} and S_H results were measured with the same setup. Therefore most of the systematic errors cancel, and this results in rather accurate ratios with $2\sigma \sim 2.8\%$. Many authors have extracted S_{He}/S_H ratios by comparing S_{He} and S_H measured by different groups. In this case, two sets of systematic errors are superimposed.

E. Velocity-proportional region.

For $v < v_b Z_1^{2/3}$, it has been demonstrated experimentally (see, e.g., Hvelplund and Fastrup (1968) and Hvelplund (1971)) that as a function of Z_1 and Z_2 , the electronic stopping exhibits oscillations around the smooth curve given by Eq. (12). These oscillations may be understood in terms of the Lindhard-Scharff picture. The stopping power is proportional to the transport cross section for electrons scattered by a screened $L\mathcal{J}$ potential around the ion, and this cross section exhibits oscillations similar to the one responsible for the Ramsauer-Townsend effect encountered in the scattering of low energy electrons by atoms (Lindhard and Finnemann (1968)). Since the Z oscillations damp out at increasing energies, deviations from the $E^{1/2}$ dependence of S_e are expected, and experimental electronic-stopping cross sections are usually fitted to the convenient form $S_e = k \times E^p$, where k and p are constants. In Table IV, the present S_{He} results for the energy interval of 100-350 keV are presented in this form. (With E expressed in keV, the resulting stopping cross sections are obtained in units of 10^{-15} eV cm²/atom.)

Table IV

$S_e = kE^p$	H ₂	He	N ₂	O ₂	Ne	Ar	Kr	Xe
k	1.45	1.86	5.46	5.59	3.72	6.85	5.12	7.28
p	0.37	0.36	0.35	0.34	0.39	0.41	0.48	0.49

From Fig. 16, the argon data of Weyl (1953) for $150 \lesssim E_{He} \lesssim 300$ keV are observed to be in perfect agreement with the present results. The energy dependence of the argon and N_2 data of Ormrod (1968) for $10 \lesssim E_{He} \lesssim 100$ keV deviates from that of the present data. It has, however, also been observed in previous experiments that the exponent p may depend rather strongly on the energy interval.

§6. Conclusion

Stopping powers for hydrogen ions in the energy region 40 keV to 1 MeV and helium ions of 100 keV to 2.4 MeV have been measured in H_2 , He, N_2 , O_2 , CO_2 , Ne, Ar, Kr, and Xe with an accuracy of $\pm 2.5\%$ (2σ). While the hydrogen-stopping powers show good agreement with most other published data and with Andersen and Ziegler's empirical stopping-power tabulations, the helium-stopping powers are systematically lower than those of the Baylor group by 1-6%. Higher-order Z_1 effects greatly influence the evaluation of shell corrections. With these effects taken into account, the empirical shell corrections, extracted from the experimental proton stopping-power data, are in good agreement with Bonderup's theoretical calculations, based on the Lindhard-Scharff model, at energies as low as $0.1 \lesssim E_H \lesssim 1$ MeV. Experimental I values are extracted and are in satisfactory agreement with those given by Andersen and Ziegler. Z_1^3 correction terms to the Bethe formula have been deduced from the experimental data, and within the accuracy of the experiment, they agree with Lindhard's and Esbensen's value. From a comparison of the stopping powers for helium and hydrogen ions at the same velocity, it has been shown that S_{He}/S_H for $1.25 \lesssim v/v_0 \lesssim 5$ depends strongly on Z_2 and deviates significantly from the mean-square charge states $\langle Z^2 \rangle_{He \rightarrow Z_2}$ obtained directly in equilibrium charge-state measurements.

Acknowledgements

We are much indebted to E. Bonderup and J. Lindhard for their continuous interest in the present work and for the help and guidance received from them. The assistance extended to us from the technical staff, especially V. Toft, is gratefully acknowledged. Finally, we wish to thank A. Grandjean and I. Schmidt for their competent assistance in the preparation of the present article.

References

- H. H. Andersen, J. F. Bak, H. Knudsen, and B. R. Nielsen, *Phys. Rev. A* **16**, 1929 (1977).
- H. H. Andersen, F. Besenbacher, and H. Knudsen, *Nucl. Instrum. Methods* **149**, 121 (1978).
- H. H. Andersen, A. F. Garfinkel, C. C. Hanke, and H. Sørensen, *Kgl. Dan. Vidensk. Selsk. Mat. Fys. Medd.* **35**, No 4 (1966).
- H. H. Andersen, H. Simonsen, and H. Sørensen, *Nucl. Phys. A* **125**, 171 (1969).
- H. H. Andersen and J. F. Ziegler, *Hydrogen Stopping Powers and Ranges in All Elements* (Pergamon, N.Y., 1977).
- J. C. Ashley, R. H. Ritchie, and W. Brandt, *Phys. Rev. B* **5**, 2393 (1972) *ibid A* **8**, 2402 (1973).
- F. Besenbacher, H. H. Andersen, P. Hvelplund, and H. Knudsen (to be published) (1980).
- F. Besenbacher, J. Heinemeier, P. Hvelplund, and H. Knudsen, *Phys. Lett.* **61A**, 75 (1977).
- H. A. Bethe, *Ann. Phys.* **5**, 325 (1930).
- H. Bichsel, in: *Studies in Penetration of Charged Particles in Matter*, NAS-NRC Publ. 1133 (1964) p. 17.
- F. Bloch, *Ann. Phys.* **16**, 285 (1933).
- F. Bloch, *Z. Phys.* **81**, 363 (1933).
- N. Bohr, *Kgl. Danske Vidensk. Selskab, Mat. Fys. Medd.* **18**, No. 8 (1948).
- E. Bonderup, *Kgl. Danske Vidensk. Selsk. Mat. Fys. Medd.* **35**, No 17 (1967).
- E. Bonderup and P. Hvelplund, *Phys. Rev. A* **4**, 562 (1971).
- E. Bonderup and J. Lindhard (1967) (private communication).
- P. D. Bourland, W. K. Chu, and D. Powers, *Phys. Rev. B* **3**, 3625 (1971).
- J. E. Brolley and F. L. Ribe, *Phys. Rev.* **98**, 1112 (1955).
- A. B. Chilton, J. N. Cooper, and J. C. Harris, *Phys. Rev.* **93**, 413 (1954).
- W. K. Chu and D. Powers, *Phys. Rev. B* **4**, 10 (1971).
- W. K. Chu and D. Powers, *Phys. Lett.* **40A**, 23 (1972).
- V. Dose and G. Sele, *Z. Phys. A* **272**, 237 (1975).
- H. Ebsensen, Thesis, University of Aarhus (unpublished) (1977).
- B. Fastrup, A. Borup, and P. Hvelplund, *Can. J. Phys.* **46**, 489 (1968).
- B. Fastrup, P. Hvelplund, and C. A. Sautter, *Kgl. Dan. Vidensk. Selsk. Mat. Fys. Medd.* **35**, No 10 (1966).
- C. C. Hanke and H. Bichsel, *Kgl. Dan. Vidensk. Selsk. Mat.-Fys. Medd.* **38**, No 3 (1970).
- H. H. Heckman, in: *Penetration of Charged Particles in Matter: A Symposium: NAS-NRC Publication*, Washington, D.C. (1970).
- J. Heinemeier, P. Hvelplund, and F. R. Simpson, *J. Phys. B* **8**, 1880 (1975).
- U. Hoyer and H. Wäffler, *Z. Naturforsch.* **26a**, 592 (1971).
- P. Hvelplund, Thesis, University of Aarhus (unpublished) (1968).
- P. Hvelplund, *Kgl. Dan. Vidensk. Selsk. Mat.-Fys. Medd.* **38**, No 4 (1971).
- P. Hvelplund and B. Fastrup, *Phys. Rev.* **165**, 408 (1968).
- M. Inokuti, T. Baer, and J. L. Dehmer, *Phys. Rev. A* **17**, 1229 (1978).
- J. D. Jackson and R. L. McCarthy, *Phys. Rev. B* **6**, 4131 (1972).
- G. D. Kerr, L. M. Hairr, N. Underwood, and A. W. Walker, *Health Phys.* **12** 1475 (1966).
- H. Knudsen, F. Besenbacher, J. Heinemeier, and P. Hvelplund, *Phys. Rev. A* **13**, 2095 (1976).
- R. A. Langley, *Phys. Rev. B* **12**, 3575 (1975).
- J. Lindhard, *Kgl. Dan. Vidensk. Selsk. Mat.-Fys. Medd.* **28**, No 8 (1954).
- J. Lindhard, in: *Studies in Penetration of Charged Particles in Matter*, NAS-NRC Publ. 1133 (1964) p. 1.
- J. Lindhard, *Nucl. Instr. Methods* **132**, 1 (1976).

- J. Lindhard and J. Finnemann, private communication and J. Finnemann, Thesis, University of Aarhus (unpublished) (1968).
- J. Lindhard and M. Scharff, Kgl. Dan. Vidensk. Selsk. Mat.-Fys. Medd. 27, No 15 (1953).
- J. Lindhard and M. Scharff, in: *Report on Conference on Penetration of Atomic Particles*. Gatlinburg, 1958, NAS-NRC Publ. 752, (1960) p. 49.
- J. Lindhard and M. Scharff, Phys. Rev. 124, 128 (1961).
- J. Lindhard and Aa. Winther, Kgl. Danske Vidensk. Selsk. Mat.-Fys. Medd. 34, No 4 (1964).
- J. H. Ormrod, Can. J. Phys. 46, 497 (1968).
- R. B. J. Palmer, Proc. Phys. Soc. Lond. 87, 681 (1966).
- J. T. Park and E. J. Zimmermann, Phys. Rev. 131, 1611 (1963).
- J. A. Phillips, Phys. Rev. 90, 532 (1953).
- L. I. Pivovarov, V. W. Tubaevev, and M. T. Noikov, Zh. Eksp. Teor. Fiz. 41, 26 (1961) [Sov. Phys.-JETP 14, 20 (1962)].
- J. J. Ramirez, R. M. Prior, J. B. Swint, A. R. Quinton, and R. A. Blue, Phys. Rev. 179, 310 (1969).
- H. K. Reynolds, D. N. F. Dunbar, W. A. Wenzel, and W. Whaling, Phys. Rev. 92, 742 (1953).
- C. C. Rousseau, W. K. Chu, and D. Powers, Phys. Rev. A 4, 1066 (1971).
- E. Rotondi, Radiat. Res. 33, 1 (1968).
- C. A. Sautter and E. J. Zimmermann, Phys. Rev. A 140, 490 (1965).
- P. Sigmund, Phys. Rev. A 14, 996 (1976).
- E. P. Steinberg, S. B. Kaufman, B. D. Wilkins, C. E. Gross, and M. J. Fluss, Nucl. Instr. Methods 99, 309 (1972).
- J. B. Swint, R. M. Prior, and J. J. Ramirez, Nucl. Instr. Methods 80, 134 (1970).
- J. H. Thorngate, ORNL-TM-5165 (1976).
- P. V. Vavilov, Zh. Eksp. Teor. Fiz. 32, 920 (1957) [Sov. Phys.-JETP 5, 749 (1957)].
- D. Ward, J. S. Foster, H. R. Andrews, I. V. Mitchell, G. C. Ball, W. G. Davies, and G. J. Costa, AECL-5313 (1976).
- F. Wenger, R. P. Gardner, and K. Verghese, Health Phys. 25, 67 (1973).
- P. K. Weyl, Phys. Rev. 91, 289 (1953).
- P. K. Weyl and M. T. Burgy (1957) (unpublished); their results are quoted by D. I. Porat and K. Ramavataram, Proc. Phys. Soc. Lond. A 252, 394 (1960).
- R. L. Wolke, W. N. Bishop, E. Eichler, N. R. Johnson, and G. D. O'Kelley, Phys. Rev. 129, 2591 (1963).
- J. F. Ziegler, *Helium Stopping Powers and Ranges in All Elements* (Pergamon, N.Y., 1978).
- J. F. Ziegler and W. K. Chu, Atomic Data and Nuclear Data Tables 13, 463 (1974).

Indleveret til Selskabet november 1978.

Færdig fra trykkeriet august 1979.

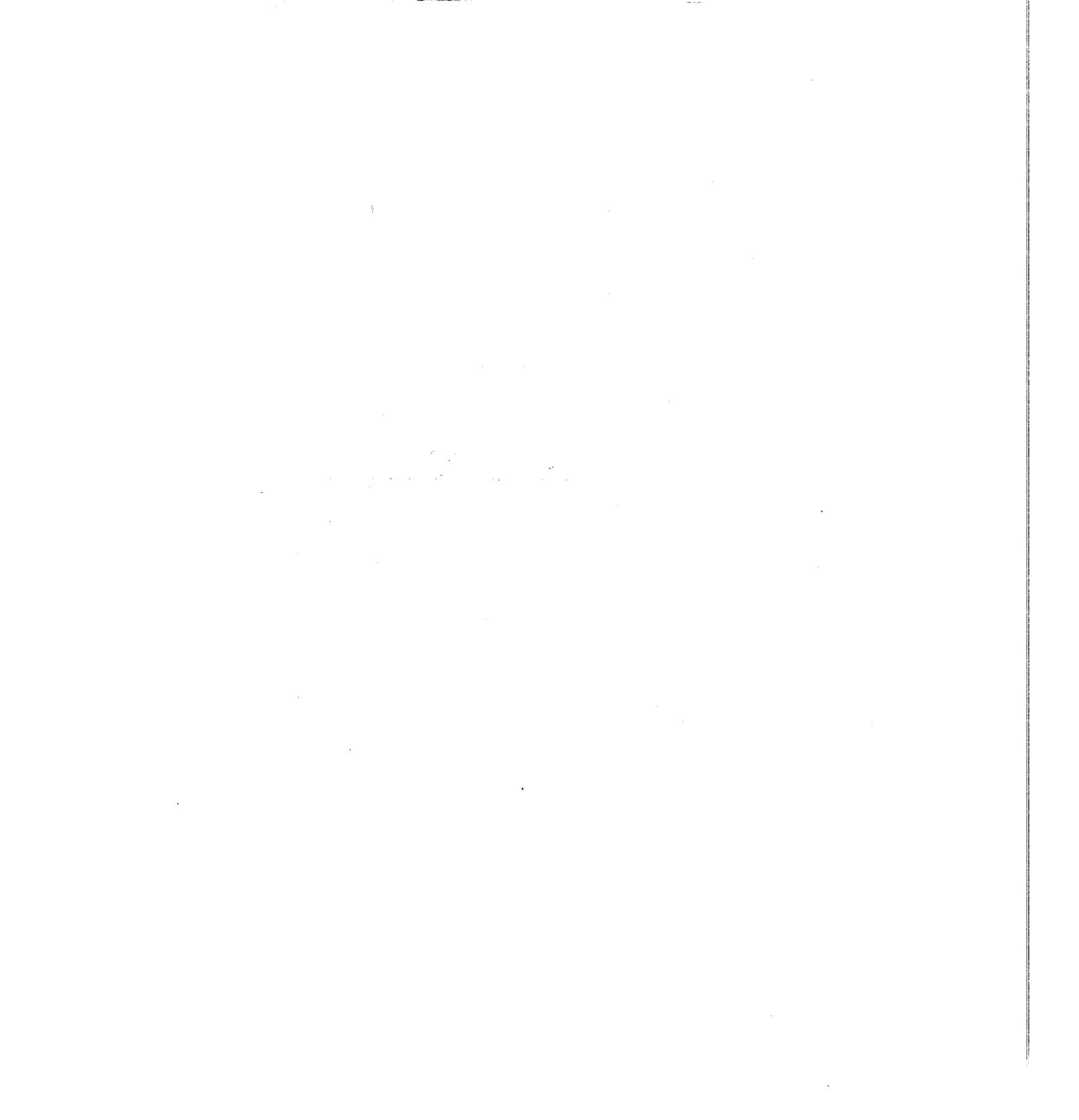


Table XIII. Experimental and calculated spectroscopic constants of the Cr_2 molecule.

Method	Basis set Type; primitive-contracted	State	Spectroscopic constants		
			r_e (a.u.)	ω_e (cm^{-1})	D_e (eV)
Experimental			3.184 ³⁾	~ 470 ²⁾	1.56 ¹⁾
All electron ab initio calculations					
MCSCF(APSG*)	GTO; (13s,7p,5d)-(6s,3p,2d)		3.59		~ 4
Restricted HF	GTO; (14s,11p,5d)-(8s,6p,2d)	$1\Gamma_g^+$	2.95	750	
GVB(1516 functions)	GTO; valence double zeta	$1\Gamma_g^+$	6.14	70	0.13
GVB(6064 functions)	as above	$2\Gamma_g^+$	5.78	110	0.35
MCSCF(APSG*)	GTO; (14s,11p,6d)-(8s,6p,3d) + bond centre functions s,p,d	$1\Gamma_g^+$	3.50	355	-6.76
CI(3196×3196)	as above	$2\Gamma_g^+$	6.56	58	-0.01
MCSCF(3196 functions)	as above	$1\Gamma_g^+$	5.93	92	0.14
CI(3520×3520)	GTO; (11s,8p,5d)-(5s,4p,3d) + bond centre functions s,p,d	$1\Gamma_g^+$	3.27	396	
MCSCF(3088 functions)	GTO; (14s,11p,6d,3f)-(8s,6p,4d,2f)				-1.4
Restricted HF	STO; (9s,6p,4d,3f,2g)		2.784	1181	-19.1
X α type calculations					
Local spin density		$13\Gamma_g^+$	6.92	55	0.20
Local spin density		$11\Gamma_u^+$	5.37	200	-0.02
X α -LCAO	GTO; (10s,6p,4d) + bond centre functions s,p,d	$1\Gamma_g^+$	5.20	110	1.00
X α -LCAO	as above	$13\Gamma_g^+$	6.92	55	0.20
Local spin density		$1\Gamma_g^+$	3.21	~ 450	1.80
Pseudopotential local spin density		$1\Gamma_g^+$	3.21	470	2.8
X α -LCAO; $\alpha_0^{1/2}$ -potential	GTO; (14s,11p,7d)-(9s,8p,5d)		4.01	106	1.4
X α -LCAO;corr. potential	as above		3.72	423	5.0

¹⁾ Antisymmetric products of strongly orthogonal geminals.

²⁾ D $^{\circ}$, Kant and Strauss 1966. - 2) Bondybey and English 1983. - 3) Efremov, Somoilova, and Gurvich 1978. - 4) Wood, Doran, Hillier, and Guest 1980. - 5) Wolf and Schmidtke 1980. - 6) Goodgame and Goddard III 1981. - 7) Atha and Hillier 1982. - 8) Çok and Hall 1983. - 9) Walch, Bauschlicher Jr., Roos, and Nelin 1983. - 10) McLean and Liu 1983. - 11) Harris and Jones 1979. - 12) Dunlap 1983. - 13) Delley, Freeman, and Ellis 1983. - 14) Bernholc and Holzwarth 1983. - 15) Baykara, McMaster, and Salahub 1984.

romagnetic coupling of localized spins of 3 on each of the Cr atoms. Thus, not even the 4s orbitals form a bonding molecular orbital at the calculated equilibrium distance. From their calculations, they determined the value of the exchange coupling constant as -93 cm^{-1} at the calculated equilibrium distance.

Kok and Hall 1983 have performed calculations that reproduce quite well the experimentally determined values of the equilibrium distance and the vibrational frequency for Cr_2 . However, they do not report any value for the dissociation energy, and without doubt, the Cr_2 molecule is unbound in their description. The nice results obtained in their calculations, most likely, stem from basis set superposition errors combined with an inadequacy of their CI wave function to allow the molecule to dissociate properly into two Cr atoms in the ${}^7\text{S}$ ground term. The wave function determined by Kok and Hall arises from the natural orbital configuration $(3d\sigma_g)^{1.78} (3d\pi_u)^{3.62} (3d\delta_g)^{3.20} (3d\delta_u)^{0.80} (3d\pi_g)^{0.38} (3d\sigma_u)^{0.22} (4s\sigma_g)^{1.92} (4s\sigma_u)^{0.08}$ at the internuclear distance 3.27 a.u. This configuration is very similar to those derived by Walch et al. 1983 in MCSCF calculations with 3088 functions. Thus, at the internuclear distance 3.0 a.u. their wave function has the natural orbital configuration $(3d\sigma_g)^{1.83} (3d\pi_u)^{3.70} (3d\delta_g)^{3.36} (3d\delta_u)^{0.64} (3d\pi_g)^{0.30} (3d\sigma_u)^{0.16} (4s\sigma_g)^{1.90} (4s\sigma_u)^{0.11}$ and at 3.5 a.u. $(3d\sigma_g)^{1.60} (3d\pi_u)^{3.32} (3d\delta_g)^{2.58} (3d\delta_u)^{1.42} (3d\pi_g)^{0.68} (3d\sigma_u)^{0.39} (4s\sigma_g)^{1.85} (4s\sigma_u)^{0.15}$.

This indicates the similarity of the descriptions of the Cr_2 molecule in Kok and Hall's and in Walch et al.'s calculations, but in Walch et al.'s calculations the Cr_2 molecule is unbound by 1.4 eV.

The $X\alpha$ type calculations also result in diversified values of the spectroscopic constants as is seen from Table XIII.

It is recognized that in spite of the vast amount of theoretical work performed for the Cr_2 molecule no clear picture of the chemical bond in this molecule has emerged. It is possible, however, that internal CI calculations within the 3d and 4s shells, comparable to the calculations done by Goodgame and Goddard 1981 and by Atha and Hillier 1982, will provide a reasonable description of the nature of the chemical bond in the Cr_2 molecule when performed at the experimental internuclear distance.

This would be in agreement with the results obtained for the «d electron rich» transition metal dimers. For the molecules Cu_2 and Ni_2 the calculated spectroscopic constants derived from extensive CI calculations are in reasonable agreement with the experimental values. However, these calculations did not alter the qualitative description of the chemical

XIV. Experimental and calculated spectroscopic constants of the Mo₂ molecule.

Method	Basis set Type; primitive-contracted	State	Spectroscopic constants		
			r _e (a.u.)	ω _e (cm ⁻¹)	D _e (eV)
Experimental			3.666 ³⁾	477.1 ²⁾	4.18 ¹⁾
All electron ab initio					
MCSCF (APSG*)	GTO; (27s, 13p, 9d) - (8s, 4p, 3d)		3.97		~ -1
CI (3212×3212)	GTO; (27s, 13p, 9d) - (9s, 5p, 4d)		3.89	392	
CI (3212×3212)	GTO; (27s, 13p, 9d) - (9s, 5p, 4d) + bond centre functions s, p, d		3.72	475	
GVB (6064 functions)	GTO; valence double zeta + 1f	¹ Γ _g ⁺	3.72	455	1.41
GVB	as above	³ Γ _u ⁺	3.91	325	0.60
Restricted HF	GTO; (18s, 11p, 9d) - (12s, 5p, 4d) + bond centre functions 2s, 2p, 1d	¹ Γ _g ⁺	3.48	699	-9.35
MCSCF (APSG*)	as above	¹ Γ _g ⁺	3.82	392	-3.10
CI (3196×3196)	as above	¹ Γ _g ⁺	3.80	388	0.86
Restricted HF	STO; (11s, 8p, 6d, 3f, 2g)	¹ Γ _g ⁺	3.400	717	-18.1
Xα type calculations					
Xα-SW		¹ Γ _g ⁺	4.3		
Local spin density		¹ Γ _g ⁺	3.69	520	4.35
Pseudopotential local spin density		¹ Γ _g ⁺	3.97	360	4.2
Xα-LCAO; α ¹ Δ-potential	GTO; (17s, 11p, 9d) - (14s, 9p, 7d)		~5.1	92	0.4
Xα-LCAO; corr. potential	as above		3.17	441	2.6

* Antisymmetric products of strongly orthogonal geminals.

1) D₀^o, Gupta, Atkins, and Gingerich 1978. - 2) Efremov, Somoilova, Kozhukhovskiy, and Gurvich 1978. - 3) Hopkins, Langridge-Smith, Morse, and Smalley 1983. - 4) Wood, Doran, Hillier, and Guest 1980. - 5) Bursten, Cotton, and Hall 1980. - 6) Goodgame and Goddard III 1982. - 7) Atha, Hillier, and Guest 1980; Atha and Hillier 1982. - 8) McLean and Liu 1983. - 9) Norman, Kolari, Gray, and Trogler 1977. - 10) Delley, Freeman, and Ellis 1983. - 11) Bernholc and Holzwarth 1983. - 12) Baykara, McMaster, and Salahub 1984.

bond derived from internal CI calculations within the 3d and 4s shells. A similar relationship for the »d electron deficient« transition metal dimers still needs to be proven.

Mo₂. Theoretical investigations of the Mo₂ molecule are up against similar difficulties as those encountered in the Cr₂ molecule. Like the Cr

atom, the ground term of the Mo atom is ${}^7S(4d)^5(5s)^1$ well separated from the higher lying terms. Intuitively, it is to be expected that the ground state of Mo_2 is a ${}^1\Sigma_g^+$ state, which should be well described in the molecular orbital picture by a one determinant wave function with all valence bonding orbitals fully occupied: $(4d\sigma_g)^2(4d\pi_u)^4(4d\delta_g)^4(5s\sigma_g)^2$. Furthermore, such a description is supported by the experimental findings of the very short bond distance, 3.666 a.u., and the high vibrational frequency, 477.1 cm^{-1} .

However, the data presented in Table XIV reveal that the accurate ab initio calculations based on the above assumptions lead to too short bond distances, too high vibrational frequencies and, more disturbing, the molecule is unbound by 9 eV in the calculations by Atha et al. 1980, 1982, and by 18 eV in McLean and Liu's 1983 calculations, which presumably are very close to the HF limit.

Contrary to the findings for Cr_2 , however, reasonable spectroscopic constants have been obtained in the calculations going beyond HF. Basically, the methods used in these calculations are very much alike, and they also lead to spectroscopic data in quantitative agreement.

Bursten et al. 1980 and Atha et al. 1980, 1982 have optimized their molecular orbitals in MCSCF type calculations, which included selected configurations within the space spanned by the 4d and 5s valence orbitals. Both groups utilized the optimized molecular orbitals in internal CI calculations within the 4d and 5s shells, but including only restricted excitations. Bursten et al. included all single and double excitations from the 64 determinants corresponding to perfect pairing in the GVB method. Atha et al. included all configurations in which the number of electrons in the $4d\sigma$ as well as in the $5s\sigma$ added up to 2 while those in the $4d\pi$ and in the $4d\delta$ added up to 4. These restrictions reduced the dimension of the CI matrix to 3212 and 3196, respectively, as compared to app. 35 000 for a full internal CI calculation within the 4d and 5s shells.

At the internuclear distance 3.78 a.u. the wavefunction derived by Bursten et al. has the configuration $(4d\sigma_g)^{1.92}(4d\pi_u)^{3.78}(4d\delta_g)^{3.42}(4d\delta_u)^{0.58}(4d\pi_g)^{0.22}(4d\sigma_u)^{0.08}(5s\sigma_g)^{1.88}(5s\sigma_u)^{0.12}$ and that of Atha et al. $(4d\sigma_g)^{1.86}(4d\pi_u)^{3.76}(4d\delta_g)^{3.33}(4d\delta_u)^{0.67}(4d\pi_g)^{0.24}(4d\sigma_u)^{0.14}(5s\sigma_g)^{1.87}(5s\sigma_u)^{0.13}$.

The similarity of the configurations indicates that the wave function determined by Bursten et al. does describe a bound molecule with a dissociation energy similar to that obtained in Atha et al.'s work. The improvement of the spectroscopic data achieved by Bursten et al. by

inclusion of bond centre functions in their basis set probably reflects larger basis set superposition errors.

The calculations by Goodgame and Goddard 1982 are more elaborate than those by Bursten et al. and by Atha et al., but not much different in principle. Their wave functions are constructed by assigning the 4d and 5s electrons pairwise to either doubly occupied bonding orbitals, doubly occupied antibonding orbitals or one in a bonding and the other in the corresponding antibonding orbital. This gives rise to $3^6=729$ spatial configurations, or 1516 spin eigenfunctions for the $^1\Sigma_g^+$ state. On top of this they include excitations from the 5s orbitals into the 5p σ and the 5p π orbitals. The resulting wavefunctions consist of 6064 spin eigenfunctions and they are optimized in MCSCF type calculations. The improved spectroscopic constants derived in Goodgame and Goddard's work as compared to those derived by Bursten et al. and by Atha et al. are partly due to the better optimization of the orbitals and partly due to inclusion of correlation of the 5s electrons.

It is gratifying to experience that acceptable spectroscopic constants for the Mo₂ molecule can be derived from conceptually simple wave functions including external correlation only of the 5s electrons. In addition, this certainly implies that the method we have used, that is full internal CI based on molecular orbitals optimized in HHF like calculations, to describe the electronic structure of the »d electron rich« transition metal dimers is appropriate, indeed.

V₂. Recently, Langridge-Smith et al. 1984 have investigated the V₂ molecule spectroscopically in the gas phase. Their results indicate that the molecule has a $^3\Sigma_g^-$ ground state. Furthermore, they determined the equilibrium distance as 3.34 a.u., and the vibrational frequency as 535 cm⁻¹. From evidence of predissociation they derived the dissociation energy as 1.85 eV, and this is considerably less than the value, 2.47 eV, derived from high temperature mass spectrometric measurements.

The first theoretical investigation of the V₂ molecule has been performed by Harris and Jones 1979 using a local spin density method. Their calculations resulted in a $^9\Sigma_u^-$ ground state with the configuration $(3d\sigma_g)^1 (3d\pi_u)^2 (3d\delta_g)^2 (3d\delta_u)^2 (3d\sigma_u)^1 (4s\sigma_g)^2$, but they found numerous other states with similar energies. The calculated spectroscopic constants for the $^9\Sigma_u^-$ state were derived as $r_c=5.01$ a.u., $\omega_c=230$ cm⁻¹, and $D_e=1.00$ eV.

Apparently the experimental results by Langridge-Smith et al. inspired Walch et al. 1983 to perform theoretical investigations of a $^3\Sigma_g^-$

state of V_2 . They performed MCSCF type calculations using two different basis sets consisting of Gaussian type functions. Their smaller basis set consisted of (14s, 11p, 6d) primitive functions which they contracted to (8s, 6p, 4d). The larger basis set was derived from the smaller by addition of 3 primitive f functions contracted to 2.

The ground term of the V atom is ${}^6D(3d)^3(4s)^2$, but the chemical bond in V_2 probably arises from two V atoms interacting in the ${}^4F(3d)^4(4s)^1$ term that is only 0.25 eV above the ground term. In this case, a simple molecular orbital picture suggests that the V_2 molecule has a ${}^3\Sigma_g^-$ ground state with the configuration $(3d\sigma_g)^2 (3d\pi_u)^4 (3d\delta_g)^2 (4s\sigma_g)^2$.

In Walch et al.'s MCSCF calculations they constrained the configurations included to those having 4 valence electrons in orbitals of σ symmetry, 2 valence electrons in π_x orbitals, 2 in π_y orbitals, 1 in δ_{xy} , and 1 in $\delta_{x^2-y^2}$. The type of calculations Walch et al. have performed especially takes into account the left-right correlation of the 3d and the 4s electrons.

In the calculations with the smaller basis set the V_2 molecule was found to be bound relative to two V atoms in the ${}^4F(3d)^4(4s)^1$ term, but not relative to two ground term V atoms. This is, however, achieved in the calculation with the larger basis set. Presumably the two sets of calculations result in very similar descriptions of the chemical bond, but this cannot be judged from the data given.

The spectroscopic constants derived from Walch et al.'s calculations with the larger basis set are in good agreement with the experimental values as to the equilibrium distance and the vibrational frequency. The calculated equilibrium distance is 3.34 a.u., the vibrational frequency is 593.6 cm^{-1} , but the calculated dissociation energy is only 0.33 eV. At the internuclear distance 3.00 a.u. the wave function has the configuration $(3d\sigma_g)^{1.83} (3d\pi_u)^{3.80} (3d\delta_g)^{1.86} (3d\delta_u)^{0.14} (3d\pi_g)^{0.20} (3d\sigma_u)^{0.15} (4s\sigma_g)^{1.93} (4s\sigma_u)^{0.09}$ and at 3.50 a.u. $(3d\sigma_g)^{1.81} (3d\pi_u)^{3.70} (3d\delta_g)^{1.74} (3d\delta_u)^{0.26} (3d\pi_g)^{0.30} (3d\sigma_u)^{0.18} (4s\sigma_g)^{1.93} (4s\sigma_u)^{0.07}$. These configurations indicate considerable participation of the 3d electrons in the bonding.

Thus, as in the case of Mo_2 , spectroscopic constants for the V_2 molecule in reasonable agreement with the experimental values have been obtained from conceptually simple wavefunctions.

Nb_2 . From high temperature mass spectrometric measurements Gupta and Gingerich 1979 determined the dissociation energy of the Nb_2

molecule to 5.21 eV. Recently Cotton and Shim 1985 have performed a theoretical investigation of the molecule using ab initio methods.

The ground term of the Nb atom is ${}^6D(4d)^4(5s)^1$, and it is expected that interaction between two Nb atoms in their ground term will result in formation of a stable molecule.

The numerous HF and HHF calculations we have performed at the internuclear distance 5.402 a.u. only revealed two configurations which are associated with energies below the sum of the energies of the separated atoms. These configurations are $(4d\sigma_g)^1 (4d\pi_u)^2 (4d\delta_g)^1 (4d\delta_u)^1 (4d\pi_g)^2 (4d\sigma_u)^1 (5s\sigma_g)^2$ and $(4d\sigma_g)^{1.0} (4d\pi_u)^{1.5} (4d\delta_g)^{1.5} (4d\delta_u)^{1.5} (4d\pi_g)^{1.5} (4d\sigma_u)^{1.0} (5s\sigma_g)^{2.0}$, respectively. The molecular orbitals optimized for the latter configuration have been utilized in CI calculations that allowed full reorganization within the 4d shells.

Fig. 11 shows all 75 low-lying electronic states of the Nb₂ molecule at the internuclear distance 5.402 a.u., which is the nearest neighbour distance in bulk Nb. It is noted that the six lowest lying states are separated from the dense manifold of higher lying states by an energy gap of app. 0.1 eV. This is of the same order of magnitude as the energy gap separating the ground state from the higher lying states in the Ru₂ molecule. The ground state of the Nb₂ molecule is a ${}^1\Sigma_g^+$ state with the configuration $(4d\sigma_g)^{1.34} (4d\pi_u)^{2.69} (4d\delta_g)^{1.11} (4d\delta_u)^{0.91} (4d\pi_g)^{1.28} (4d\sigma_u)^{0.67} (5s\sigma_g)^{2.00}$ at the internuclear distance 5.402 a.u. Two other states, ${}^1\Gamma_g$ and ${}^1\Sigma_u^-$, lying very close to the ground state arise from almost identical configurations. The lowest lying ${}^3\Sigma_g^-$ state has the configuration $(4d\sigma_g)^{1.27} (4d\pi_u)^{2.62} (4d\delta_g)^{1.09} (4d\delta_u)^{0.93} (4d\pi_g)^{1.36} (4d\sigma_u)^{0.73} (5s\sigma_g)^{2.00}$, and this is almost identical to the configurations of the lowest lying ${}^3\Sigma_u^+$ and ${}^3\Gamma_u$ states. Thus, the six lowest lying states reveal substantial participation of the 4d electrons in the bonding. For the lowest lying singlet states, the excess of electrons in bonding 4d orbitals relative to antibonding 4d orbitals is 2.28, whereas it has decreased to 1.96 electrons for the lowest lying triplet states. Judging from Mulliken population analyses, the principal bonding orbital is the delocalized $5s\sigma_g$ molecular orbital. At 5.402 a.u. this orbital contributes 0.69 out of a total overlap population of 0.81.

The equilibrium distances derived from the calculated potential energy curves for the 75 low-lying states are scattered over app. 1 a.u., ranging from 5.6-6.6 a.u. This is in contrast to the findings for the »d electron rich« transition metal dimers. Thus, for the Pd₂ molecule the corresponding spread of the equilibrium distances is only 0.26 a.u. Due to the

molecules V_2 , Cr_2 , and Mo_2 we also expect that Nb_2 has a short bond distance. The change in equilibrium distance in going from Cr_2 to Mo_2 is identical with that of going from Cu_2 to Ag_2 . If we invoke a similar relationship when going from the V_2 to the Nb_2 molecule we should expect an equilibrium distance of only 3.81 a.u. in Nb_2 . This implies, by analogy with our findings for the Ag_2 molecule, that the calculated equilibrium distance for Nb_2 should be app. 4.36 a.u. The calculated equilibrium distance, however, is app. 5.6 a.u.

We expect that the major reason for the shoulders in the potential energy curves as well as for the large discrepancy between the calculated and the expected equilibrium distance is due to the lack of f functions in the basis set. Such functions are required to account properly for the 4d electrons, because they are deeply involved in the bonding.

Based on the simple molecular orbital picture the Nb_2 molecule should have a $^3\Sigma_g^-$ ground state with the configuration $(4d\sigma_g)^2 (4d\pi_u)^4 (4d\delta_g)^2 (5s\sigma_g)^2$. HF calculations reveal that this state does not describe a bound molecule, but it should be noted that the total number of $d\sigma$ and $d\delta$ electrons each add up to app. 2 for the $^1\Sigma_g^+$ ground state, just as the $d\pi$ electrons add up to app. 4. Walch et al. 1983 imposed such constraints onto their wavefunctions for the V_2 molecule. In our work no such constraints have been imposed, and therefore the number of electrons associated with each symmetry type is a genuine result of the calculations performed. This implies that the molecular orbital picture offers a useful guidance in determining the ground state of the Nb_2 molecule. A similar result has not been found for the »d electron rich« transition metal dimers, and it still needs to be investigated, if it is valid also for other »d electron deficient« transition metal dimers.

Sc_2 . The Sc_2 molecule has been investigated by high temperature mass spectrometry by Verhaegen et al. 1964. However, the dissociation energy derived from the mass spectrometric data appears unreliable, since the original value 1.12 eV derived by Verhaegen et al. later on, without justification, has been changed to 1.65 eV as quoted in Table XV. Lately, the Sc_2 molecule has also been studied by matrix isolation techniques. This has resulted in determination of the ground state vibrational frequency as 238.91 cm^{-1} , and from the ESR studies by Knight et al. 1983 the symmetry of the ground state has been derived as $^5\Sigma$.

The first theoretical investigation of the Sc_2 molecule is a local spin density calculation performed by Harris and Jones 1979. Their calculations resulted in a $^5\Sigma_u^-(3d\sigma_g)^1(3d\pi_u)^2(3d\sigma_u)^1(4s\sigma_g)^2$ ground state with an

equilibrium distance of 5.10 a.u., a vibrational frequency of 200 cm^{-1} , and a dissociation energy of 1.80 eV. Furthermore, they also have reported the spectroscopic constants for a ${}^3\Sigma_g^- (3d\pi_u)^2 (4s\sigma_g)^2 (4s\sigma_u)^2$ state as $r_e=6.15$ a.u., $\omega_e=235\text{ cm}^{-1}$, and $D_e=1.00$ eV. Harris and Jones 1978 have determined the discrepancies in the local spin density method between the calculated and the experimental energy splittings of atomic states originating from different orbital configurations. Applying this correction to the dissociation energy for the ${}^5\Sigma_u^-$ state of Sc_2 reduces it to app. 0.4 eV. This is, however, still in disagreement with the results obtained in HF calculations by Wood et al. 1980. They found the ${}^5\Sigma_u^- (3d\sigma_g)^1 (3d\pi_u)^2 (3d\sigma_u)^1 (4s\sigma_g)^2$ state unbound by app. 2 eV relative to the atomic limit ${}^2D(3d)^1(4s)^2+{}^4F(3d)^2(4s)^1$. At least for the transition metal dimers, this kind of relationship between dissociation energies derived in HF and $X\alpha$ type calculations are common, and the reason for the discrepancy is presently not clear.

The most extensive theoretical investigation of the Sc_2 molecule has been carried out by Das 1982 using a pseudopotential method in conjunction with a large valence basis set consisting of Slater type functions. The conclusion of his work is that the Sc_2 molecule is a van der Waals dimer with a binding energy of app. 0.17 eV at the internuclear distance app. 9.4 a.u. His results, of course, are in sharp contrast to the available experimental data for the Sc_2 molecule.

The results by Das were essentially confirmed by Walch and Bauschlicher 1983a. They performed all electron ab initio calculations on a few selected states of the Sc_2 molecule and found that the states dissociating into two ${}^2D(3d)^1(4s)^2$ ground term Sc atoms were very weakly bound with equilibrium distances of app. 8 a.u. and dissociation energies of app. 0.06 eV. However, they also performed calculations on a ${}^5\Delta_u$ and a ${}^5\Sigma_u^-$ state that dissociate into one Sc atom in the ${}^2D(3d)^1(4s)^2$ ground term and the other in the ${}^4F(3d)^2(4s)^1$ excited term. Relative to this dissociation limit the ${}^5\Delta_u$ and the ${}^5\Sigma_u^-$ states were strongly bound by app. 0.8 eV.

The appearance of the experimental results by Knight et al. 1983 indicating the existence of a bound ${}^5\Sigma$ state for the Sc_2 molecule influenced Walch and Bauschlicher 1983b to reinvestigate the ${}^5\Sigma_u^-$ state. In their MCSCF calculations on the ${}^5\Sigma_u^-$ state they constrained the configurations included to those having a total of 4 valence electrons of σ symmetry and 2 of π symmetry. The optimized molecular orbitals were utilized in CI calculations that allowed single and double excitations from ten selected

e XV. Experimental and calculated spectroscopic constants of the Sc₂ molecule.

Method	Basis set Type; primitive-contracted	State(atomic limit)	Spectroscopic constants		
			r _e (a.u.)	ω _e (cm ⁻¹)	D _e (eV)
Experimental				238.91 ²⁾	1.65 ¹⁾
All electron ab initio calculations					
MCSCF	GTO; (13s,7p,5d)-(6s,3p,2d)	$5\Sigma_u^-(^2D(4s^23d^1)+^4F(4s^13d^2))$	4.86		0.30
CI	as above	$5\Sigma_u^-(^2D(4s^23d^1)+^4F(4s^13d^2))$	4.91		1.12
Restricted HF	GTO; (14s,11p,5d)-(8s,6p,2d)	$1\Sigma_g^+(3d\sigma_g)^2(3d\sigma_u)^2(4s\sigma_g)^2$	5.76	210	
Restricted HF	as above	$1\Sigma_g^+(4s\sigma_g)^2(3d\pi_u)^4$	4.20	360	
CI	GTO; (14s,11p,6d)-(8s,6p,4d)	$1\Sigma_g^-(^2D(4s^23d^1)+^2D(4s^23d^1))$	~8		0.045
CI	as above	$3\Sigma_g^-(^2D(4s^23d^1)+^2D(4s^23d^1))$	~8		0.045
CI	as above	$3\Sigma_u^+(^2D(4s^23d^1)+^2D(4s^23d^1))$	~8		0.029
CI	GTO; (14s,11p,6d,1f)-(6s,6p,3d,1f)	$3\Sigma_u^+(^2D(4s^23d^1)+^2D(4s^23d^1))$	~8		0.046
CI	GTO; (14s,11p,6d)-(8s,6p,4d)	$5\Delta_u(^2D(4s^23d^1)+^4F(4s^14p^13d^1))$	~7		0.8
CI	as above	$5\Sigma_u^-(^2D(4s^23d^1)+^4F(4s^14p^13d^1))$	~6.5,~7		0.2
CI	GTO; (14s,11p,6d)-(8s,6p,4d)	$5\Sigma_u^-(^2D(4s^23d^1)+^2D(4s^23d^1))$	5.27	184	0.12
CI, corrected	as above	as above	5.27	184	0.44
Pseudopotential calculations					
CI	STO; (3s,3p,4d,1f)	$1\Sigma_g^+(^2D(4s^23d^1)+^2D(4s^23d^1))$	-9.4	61	0.17
Xα type calculations					
Local spin density		$5\Sigma_u^-$	5.10	200	1.80
Local spin density		$3\Sigma_g^-$	6.15	235	1.00

1) Drowart 1967. - 2) Moskovits, DiLella and Limm 1984. - 3) Wood, Doran, Hillier and Guest 1980. - 4) Wolf and Schmidtke 1980. - 5) Walch and Bauschlicher Jr. 1983a. - 6) Walch and Bauschlicher Jr. 1983b. - 7) Das 1982. - 8) Harris and Jones 1979.

reference configurations. From their calculations they derived an equilibrium distance of 5.27 a.u., a vibrational frequency of 184 cm⁻¹, and a dissociation energy relative to two ground term atoms of 0.12 eV which they corrected to 0.44 eV by taking into account Davidson's correction, correction for errors in the asymptotic limit and also corrections for basis set insufficiencies.

Prior to the appearance of the experimental work by Knight et al. 1983, Wood et al. 1980 suggested a $5\Sigma_u^-$ ground state of the Sc₂ molecule on basis of their MCSCF and CI calculations. Their calculations, how-

ever, were not convincing by themselves due to errors in the atomic limit, and substantial superposition errors associated with the basis sets used.

Conclusion

During the past two decades great efforts have been devoted to experimental and theoretical investigations trying to elucidate the natures of the chemical bonds in small molecules composed of transition metal atoms. In the present work we have reviewed the experimental and the theoretical knowledge of the dimers of the first and second transition metal series, and in addition, we have presented some new calculational results for the molecules Ni_2 , Fe_2 , and Rh_2 . At present, it appears that a unified understanding of the transition metal dimers is emerging from the combined experimental and theoretical results, which for a long time have seemed contradictory

By now it must be considered proven that it is extremely difficult, from *ab initio* work, to derive spectroscopic constants for the transition metal dimers, which are in good agreement with the experimental values. Nevertheless, such good agreements have been achieved for a few transition metal dimers of the »d electron rich« and recently also of the »d electron deficient« types. In addition, the theoretical predictions of the many low-lying electronic states in the Ni_2 molecule have recently been confirmed by experimental work. Therefore, we are confident that *ab initio* methods in general provide physical insight into the complex chemical bonds of the transition metal dimers. In particular, it appears that adequate descriptions of the transition metal dimers are obtained in internal CI calculations involving only the valence d and s shells. Such calculations do not result in spectroscopic constants of high accuracy, but for the »d electron rich« transition metal dimers, it appears that the calculated equilibrium distances deviate systematically from the experimental values. For the »d electron deficient« transition metal dimers the theoretical data are still too scarce to reach a similar conclusion. In any case, it is far more difficult to achieve reasonable agreements for these molecules.

Due to the unfilled d shells of the transition metal atoms, the wave functions for the dimers usually consist of a large number of Slater determinants and therefore appear to be very complicated. In spite of

this, however, it has been possible to interpret the wave functions in simple physical pictures.

Thus, the chemical bonds in the »d electron rich« transition metal dimers, Cu_2 , Ni_2 , Co_2 , and Fe_2 are mainly due to the delocalized $4s\sigma_g$ molecular orbitals. The 3d electrons localize around the nuclei and interact through Heisenberg exchange couplings. This gives rise to the small energy splittings between the many low-lying electronic states. The principal bonding orbitals in the corresponding dimers of the second transition metal series are the delocalized $5s\sigma_g$ molecular orbitals. The 4d electrons in Ag_2 and Pd_2 are well localized, but in the sequence Rh_2 to Ru_2 the 4d electrons become increasingly delocalized signifying their participation in the bonding. The d electrons in the »d electron deficient« transition metal dimers are considerably delocalized and strongly involved in the bonding of the molecules. In these molecules the combined contributions to the bonding from the d electrons are comparable to the contributions from the outermost $s\sigma_g$ molecular orbitals.

The chemical bonds in the transition metal dimers cannot be described in a molecular orbital picture, but at least in the case of the Nb_2 molecule, it appears that the simple picture still has a significance.

In summary, accurate spectroscopic constants for the transition metal dimers are not obtained from ab initio calculations, but it appears that the nature of the chemical bonds in such molecules are well described by the conceptually simple wave functions resulting from internal CI calculations within the valence d and s shells.

ACKNOWLEDGEMENTS. The author is deeply indebted to Professor J. P. Dahl for scientific guidance and for sharing his inspiring and profound knowledge of the field of quantum chemistry. The author wishes to express her gratitude to Professor F. A. Cotton and Professor K. A. Gingerich for inspiring collaboration that has added many chemical aspects to the work. Furthermore, Professor K. A. Gingerich's deep interest and never ceasing encouragements are very much appreciated. Finally, the valuable discussions with Dr. Helge Johansen and Dr. Sten Rettrup over the years are greatly acknowledged.

Part of the computations have been performed at NEUCC, the Technical University of Denmark and supported financially by the Danish Natural Science Research Council. The major computations have been carried out at CSC, Texas A&M University and generously supported by Texas A&M University.

References

- F. Ahmed and E. R. Nixon, *J. Chem. Phys.* **71**, 3547 (1979).
P. M. Atha, I. H. Hillier, and M. F. Guest, *Chem. Phys. Letters* **75**, 84 (1980).
P. M. Atha and I. H. Hillier, *Mol. Phys.* **45**, 285 (1982).
C. Bachmann, J. Demuynck, and A. Veillard, *Gazz. Chim. Ital.* **108**, 389 (1978).
C. Bachmann, J. Demuynck, and A. Veillard, *Faraday Symp. Chem. Soc.* **14**, 170 (1980).
P. H. Barrett and T. K. McNab, *Phys. Rev. Letters* **25**, 1601 (1970).
H. Basch, *Faraday Symp. Chem. Soc.* **14**, 149 (1980)a.
H. Basch, D. Cohen, and S. Topiol, *Isr. J. Chem.* **19**, 233 (1980)b.
H. Basch, *J. Am. Chem. Soc.* **103**, 4657 (1981).
C. A. Baumann, R. J. Van Zee, S. V. Bhat, and W. Weltner, Jr., *J. Chem. Phys.* **78**, 190 (1983).
C. W. Bauschlicher, Jr., S. P. Walch, and P. E. M. Siegbahn, *J. Chem. Phys.* **76**, 6015 (1982).
C. W. Bauschlicher, Jr., *Chem. Phys. Letters* **97**, 204 (1983).
N. A. Baykara, B. N. McMaster, and D. R. Salahub, *Mol. Phys.* **52**, 891 (1984).
E. D. Becker and G. C. Pimentel, *J. Chem. Phys.* **25**, 224 (1956).
J. Bernholc and N. A. W. Holzwarth, *Phys. Rev. Letters* **50**, 1451 (1983).
A. Bino, F. A. Cotton, and T. R. Fethouse, *Inorg. Chem.* **18**, 2599 (1979).
N. Bohr, *Phil. Mag.* **26**, 1 (1913); *ibid* **26**, 476 (1913).
V. E. Bondybey and J. H. English, *Chem. Phys. Letters* **94**, 443 (1983).
Ø. Burrau, *Kgl. Danske Vid. Selsk. Mat.-Fys. Medd.* **7**, 14 (1927).
B. E. Bursten and F. A. Cotton, *Faraday Symp. Chem. Soc.* **14**, 180 (1980).
B. E. Bursten, F. A. Cotton, and M. B. Hall, *J. Am. Chem. Soc.* **102**, 6348 (1980).
T. A. Cellucci and E. R. Nixon, *J. Chem. Phys.* **81**, 1174 (1984).
M. B. Cingi, D. A. Clemente, and C. Foglia, *Mol. Phys.* **53**, 301 (1984).
C. R. Claydon and K. D. Carlson, *J. Chem. Phys.* **49**, 1331 (1968).
D. L. Cocke and K. A. Gingerich, *J. Chem. Phys.* **60**, 1958 (1974).
C. Cossé, M. Fouassier, T. Mejean, M. Tranquille, D. P. DiLella, and M. Moskovits, *J. Chem. Phys.* **73**, 6076 (1980).
F. A. Cotton and I. Shim, *J. Am. Chem. Soc.* **104**, 7025 (1982).
F. A. Cotton and I. Shim, *J. Phys. Chem.*, **89**, 952 (1985).
G. Das, *Chem. Phys. Letters* **86**, 482 (1982).
B. Delley, A. J. Freeman, and D. E. Ellis, *Phys. Rev. Letters* **50**, 488 (1983).
J. P. Desclaux, *Atom. Data Nucl. Data Tables* **12**, 311 (1973).
D. P. DiLella, W. Limm, R. H. Lipson, M. Moskovits, and K. V. Taylor, *J. Chem. Phys.* **77**, 5263 (1982).
P. A. M. Dirac, *Proc. Roy. Soc. A* **123**, 714 (1929).
R. N. Dixon and I. L. Robertson, *Mol. Phys.* **36**, 1099 (1978).
J. Drowart and R. E. Honig, *J. Chem. Phys.* **25**, 581 (1956).
J. Drowart and R. E. Honig, *J. Phys. Chem.* **61**, 980 (1957).
J. Drowart, in »Phase Stability in Metals and Alloys«. Edited by P. S. Rutman, J. Stringer and R. I. Jaffee (McGraw-Hill, New York 1967) pp 305-317.
B. I. Dunlap and H. L. Yu, *Chem. Phys. Letters* **73**, 525 (1980).
B. I. Dunlap, *Phys. Rev. A* **27**, 2217 (1983).

- Y. M. Efremov, A. N. Samoilova, and L. V. Gurvich, *Opt. Spektrosc.* **36**, 654 (1974).
Y. M. Efremov, A. N. Samoilova, V. B. Kozhukhovskiy, and L. V. Gurvich, *J. Mol. Spec.* **73**, 430 (1978).
A. Einstein, *Ann. Physik* **17**, 132 (1905).
V. Fock, *Zeit. Physik* **61**, 126 (1930).
T. A. Ford, H. Huber, W. Klotzbücher, E. P. Kündig, M. Moskovits, and G. A. Ozin, *J. Chim. Phys.* **66**, 524 (1977).
J. García-Prieto and O. Novaro, *Int. J. Quantum Chem.* **18**, 595 (1980).
K. A. Gingerich, *Naturwissenschaften* **54**, 43 (1967).
K. A. Gingerich and D. L. Cocke, *Chem. Commun.* 536 (1972).
K. A. Gingerich, *Curr. Top. Mat. Sci.* **6**, 345 (1980).
M. M. Goodgame and W. A. Goddard III, *J. Phys. Chem.* **85**, 215 (1981).
M. M. Goodgame and W. A. Goddard III, *Phys. Rev. Letters* **48**, 135 (1982).
D. W. Green and D. M. Gruen, *J. Chem. Phys.* **57**, 4462 (1972).
D. Guenzburger and E. M. B. Saitovitch, *Phys. Rev.* **B24**, 2368 (1981).
S. K. Gupta, R. M. Atkins, and K. A. Gingerich, *Inorg. Chem.* **17**, 3211 (1978).
S. K. Gupta and K. A. Gingerich, *J. Chem. Phys.* **70**, 5350 (1979).
J. Harris and R. O. Jones, *J. Chem. Phys.* **68**, 3316 (1978).
J. Harris and R. O. Jones, *J. Chem. Phys.* **70**, 830 (1979).
D. R. Hartree, *Proc. Cambridge Phil. Soc.* **24**, 89 (1928).
W. Heitler and F. London, *Zeit. Physik* **44**, 455 (1927).
J. B. Hopkins, P. R. R. Langridge-Smith, M. D. Morse, and R. E. Smalley, *J. Chem. Phys.* **78**, 1627 (1983).
K. P. Huber and G. Herzberg, »Molecular Spectra and Molecular Structure IV, Constants of Diatomic Molecules« (Van Nostrand Reinhold, New York 1979).
F. Hund, *Zeit. Physik* **51**, 759 (1928).
S. Huzinaga, *J. Chem. Phys.* **66**, 4245 (1977).
E. A. Hylleraas, *Zeit. Physik* **65**, 209 (1930).
P. Joyes and M. Leleyter, *J. Phys.* **B6**, 150 (1973).
A. Kant, *J. Chem. Phys.* **41**, 1872 (1964).
A. Kant and B. Strauss, *J. Chem. Phys.* **41**, 3806 (1964).
A. Kant and B. Strauss, *J. Chem. Phys.* **45**, 3161 (1966).
A. Kant, S.-S. Lin, and B. Strauss, *J. Chem. Phys.* **49**, 1983 (1968).
A. Kant and S.-S. Lin, *J. Chem. Phys.* **51**, 1644 (1969).
B. Kleman and S. Lindkvist, *Arkiv Fysik* **8**, 333 (1954).
B. Kleman and S. Lindkvist, *Arkiv Fysik* **9**, 385 (1955).
M. Klobukowski, *J. Comp. Chem.* **4**, 350 (1983).
W. E. Klotzbücher and G. A. Ozin, *Inorg. Chem.* **16**, 984 (1977).
W. E. Klotzbücher and G. A. Ozin, *Inorg. Chem.* **19**, 2848 (1980)a.
W. E. Klotzbücher and G. A. Ozin, *Inorg. Chem.* **19**, 3767 (1980)b.
L. B. Knight Jr., R. J. Van Zee, and W. Weltner, Jr., *Chem. Phys. Letters* **94**, 296 (1983).
R. A. Kok and M. B. Hall, *J. Phys. Chem.* **87**, 715 (1983).
W. Kolos, *Theoret. Chim. Acta (Berlin)* **51**, 219 (1979).
P. R. R. Langridge-Smith, M. D. Morse, G. P. Hansen, R. E. Smalley and A. J. Merer, *J. Chem. Phys.* **80**, 593 (1984).
O. Laporte and D. R. Inglis, *Phys. Rev.* **35**, 1337 (1930).
S.-S. Lin, B. Strauss, and A. Kant, *J. Chem. Phys.* **51**, 2282 (1969)a.

- S.-S. Lin and A. Kant, *J. Phys. Chem.* **73**, 2450 (1969)b.
R. L. Martin and P. J. Hay, *J. Chem. Phys.* **75**, 4539 (1981).
R. L. Martin, *J. Chem. Phys.* **78**, 5840 (1983).
A. D. McLean, *J. Chem. Phys.* **79**, 3392 (1983).
A. D. McLean and B. Liu, *Chem. Phys. Letters* **101**, 144 (1983).
T. K. McNab, H. Micklitz, and P. H. Barrett, *Phys. Rev.* **B4**, 3787 (1971).
D. L. Michalopoulos, M. E. Geusic, S. G. Hansen, D. E. Powers, and R. E. Smalley, *J. Phys. Chem.* **86**, 3914 (1982).
P. A. Montano, P. H. Barrett, and Z. Shanfield, *J. Chem. Phys.* **64**, 2896 (1976).
P. A. Montano, *Faraday Symp. Chem. Soc.* **14**, 79 (1980).
P. A. Montano and G. K. Shenoy, *Solid State Comm.* **35**, 53 (1980).
C. E. Moore, *Natl. Bur. Std. Circ. No. 467* (US Government Printing Office, Washington, D. C., 1949, 1952, 1958), vols 1, 2, and 3.
M. D. Morse, G. P. Hansen, P. R. R. Langridge-Smith, I.-S. Zheng, M. E. Geusic, D. L. Michalopoulos, and R. E. Smalley, *J. Chem. Phys.* **80**, 5400 (1984).
M. Moskovits and J. E. Hulse, *J. Chem. Phys.* **66**, 3988 (1977).
M. Moskovits and D. P. DiLella, *J. Chem. Phys.* **73**, 4917 (1980).
M. Moskovits, D. P. DiLella, and W. Limm, *J. Chem. Phys.* **80**, 626 (1984).
R. S. Mulliken, *Phys. Rev.* **32**, 186 (1928); *ibid* **32**, 761 (1928).
H. M. Nagarathna, P. A. Montano, and V. M. Naik, *J. Am. Chem. Soc.* **105**, 2938 (1983).
R. K. Nesbet, *Phys. Rev.* **A135**, 460 (1964).
J. O. Noell, M. D. Newton, P. J. Hay, R. L. Martin, and F. W. Bobrowicz, *J. Chem. Phys.* **73**, 2360 (1980).
J. G. Norman, Jr., H. J. Kolari, H. B. Gray, and W. C. Trogler, *Inorg. Chem.* **16**, 987 (1977).
G. A. Ozin, H. Huber, D. McIntosh, S. Mitchell, J. G. Norman, Jr., and L. Noodleman, *J. Am. Chem. Soc.* **101**, 3504 (1979).
M. J. Pellin, T. Foosnaes, and D. M. Gruen, *J. Chem. Phys.* **74**, 5547 (1981).
M. J. Pellin and D. M. Gruen, *J. Chem. Phys.* **79**, 5887 (1983).
M. Pelissier, *J. Chem. Phys.* **75**, 775 (1981).
V. Piacente, G. Balducci, and G. Bardi, *J. Less-Comm. Met.* **37**, 123 (1974).
M. Planck, *Ann. Physik* **4**, 553 (1901).
D. E. Powers, S. G. Hansen, M. E. Geusic, D. L. Michalopoulos, and R. E. Smalley, *J. Chem. Phys.* **78**, 2866 (1983).
H. Purdum, P. A. Montano, G. K. Shenoy, and T. Morrison, *Phys. Rev.* **B25**, 4412 (1982).
P. Pyykkö, J. G. Snijders, and E. J. Baerends, *Chem. Phys. Letters* **83**, 432 (1981).
R. C. Raffanetti, *J. Chem. Phys.* **58**, 4452 (1973).
S. J. Riley, E. K. Parks, L. G. Pobo, and S. Wexler, *J. Chem. Phys.* **79**, 2577 (1983).
J.-C. Rivoal, J. S. Emanpour, K. J. Zeringue, and M. Vala, *Chem. Phys. Letters* **92**, 313 (1982).
C. C. J. Roothaan, *Rev. Mod. Phys.* **23**, 69 (1951).
C. C. J. Roothaan, *Rev. Mod. Phys.* **32**, 179 (1960).
J. Ruamps, *Compt. rend.* **238**, 1489 (1954).
P. Schissel, *J. Chem. Phys.* **26**, 1276 (1957).

- E. Schrödinger, *Ann. Physik* **79**, 361 (1926); *ibid* **79**, 489 (1926); *ibid* **79**, 734 (1926); *ibid* **81**, 109 (1926).
- E. Schrödinger, *Phys. Rev.* **28**, 1049 (1926).
- W. Schulze, H. U. Becker, R. Minkwitz, and K. Manzel, *Chem. Phys. Letters* **55**, 59 (1978).
- I. Shim, J. P. Dahl, and H. Johansen, *Int. J. Quantum Chem.* **15**, 311 (1979).
- I. Shim, *Mol. Phys.* **39**, 185 (1980).a
- I. Shim, *Theoret. Chim. Acta (Berlin)* **54**, 113 (1980)b.
- I. Shim and K. A. Gingerich, *J. Chem. Phys.* **77**, 2490 (1982).
- I. Shim and K. A. Gingerich, *J. Chem. Phys.* **79**, 2903 (1983)a.
- I. Shim and K. A. Gingerich, *J. Chem. Phys.* **78**, 5693 (1983)b.
- I. Shim and K. A. Gingerich, *J. Chem. Phys.* **80**, 5107 (1984).
- J. C. Slater, *Phys. Rev.* **34**, 1293 (1929).
- J. C. Slater, *Phys. Rev.* **35**, 210 (1930).
- J. C. Slater, J. B. Mann, T. M. Wilson, and J. H. Wood, *Phys. Rev.* **184**, 672 (1969).
- V. I. Srdanov and D. S. Pešić, *J. Mol. Spectrosc.* **90**, 27 (1981).
- H. Tatewaki and S. Huzinaga, *J. Chem. Phys.* **72**, 399 (1980).
- H. Tatewaki, Y. Sakai, and S. Huzinaga, *J. Comp. Chem.* **2**, 96 (1981)a.
- H. Tatewaki, Y. Sakai, and S. Huzinaga, *J. Comp. Chem.* **2**, 278 (1981)b.
- T. H. Upton and W. A. Goddard III, *J. Am. Chem. Soc.* **100**, 5659 (1978).
- G. Verhaegen, S. Smoes, and J. Drowart, *J. Chem. Phys.* **40**, 239 (1964).
- T. C. de Vore, A. Ewing, H. F. Franzen, and V. Calder, *Chem. Phys. Letters* **35**, 78 (1975).
- A. J. H. Wachters, *J. Chem. Phys.* **52**, 1033 (1970).
- S. P. Walch and C. W. Bauschlicher, Jr., *Chem. Phys. Letters* **94**, 290 (1983)a.
- S. P. Walch and C. W. Bauschlicher, Jr., *J. Chem. Phys.* **79**, 3590 (1983)b.
- S. P. Walch, C. W. Bauschlicher, Jr., B. O. Roos, and C. J. Nelin, *Chem. Phys. Letters* **103**, 175 (1983).
- E. W. Whittle, D. A. Dows, and G. C. Pimentel, *J. Chem. Phys.* **22**, 1943 (1954).
- M. Witko and H. -O. Beckmann, *Mol. Phys.* **47**, 945 (1982).
- A. Wolf and H. -H. Schmidtke, *Int. J. Quantum Chem.* **18**, 1187 (1980).
- C. Wood, M. Doran, I. H. Hillier, and M. F. Guest, *Faraday Symp. Chem. Soc.* **14**, 159 (1980).
- R. J. Van Zee, C. A. Baumann, and W. Weltner, Jr., *J. Chem. Phys.* **74**, 6977 (1981).
- T. Ziegler, J. G. Snijders, and E. J. Baerends, *J. Chem. Phys.* **74**, 1271 (1981).
- T. Ziegler, in »Local Density Approximations in Quantum Chemistry and Solid State Physics«, edited by J. P. Dahl and J. Avery (Plenum, New York, 1984). p. 273.

HANS L. SKRIVER
*Crystal Structure
from One-Electron Theory*

ABSTRACT. We have studied the crystal structures of all the 3d, 4d, and 5d transition metals at zero pressure and temperature by means of the LMTO method and Andersen's force theorem. We find that, although the structural energy differences seem to be overestimated by the theory, the predicted crystal structures are in accord with experiment in all cases except Au. In addition we have investigated the effect of pressure upon the alkali metals (Li, Na, Rb, Cs) and selected lanthanide metals (La, Ce, Lu) and actinide metals (Th, Pa). In these cases the theory gives accurate predictions of the stability of the close-packed structures but is found to be less accurate for open structures such as α -U.

Risø National Laboratory, DK-4000 Roskilde, Denmark

1. Introduction

Many of the characteristic properties of the metallic elements are a consequence of their ability at normal temperature and pressure to form crystals in which the metal atoms are arranged in a regular pattern which repeats itself throughout the interior of the crystal. These crystals are the microscopic building blocks of all the pieces of metal which we encounter around us, and it is therefore of great importance to investigate their basic properties both experimentally and theoretically. The hope is of course that by isolating and understanding the factors that govern the stability of the crystal structures found in nature one may eventually be able to design metals with specified properties.

The crystal structures of solid state materials are established by X-ray diffraction experiments, and the results for the elemental metals are compiled in Fig. 1. It turns out that the variety of crystal structures which the metallic elements take on is limited to essentially the five types shown on Fig. 2, and that four of these five structures are so-called close-packed structures. The term close-packed refers to the fact that the fcc, hcp, dhcp, and Sm-type structures can be derived from stacking hexagonal

0.0 - 0.8

Li hcp	Be hcp												
Na hcp	Mg hcp	1	2	3.5	5	6	7	8.5	9.5	10	n_d		
K bcc	Ca fcc	Sc hcp	Ti hcp	V bcc	Cr bcc	Mn (bcc)	Fe bcc	Co hcp	Ni fcc	Cu fcc	Zn hcp		
Rb bcc	Sr fcc	Y hcp	Zr hcp	Nb bcc	Mo bcc	Tc hcp	Ru hcp	Rh fcc	Pd fcc	Ag fcc	Cd hcp		
Cs bcc	Ba bcc	Lu hcp	Hf hcp	Ta bcc	W bcc	Re hcp	Os hcp	Ir fcc	Pt fcc	Au fcc	Hg (fcc)		
Fr bcc	Ra bcc												

La dhcp	Ce fcc	Pr dhcp	Nd dhcp	Pm dhcp	Sm Sm-t	Eu bcc	Gd hcp	Tb hcp	Dy hcp	Ho hcp	Er hcp	Tm hcp	Yb hcp
Th fcc	Pa bct	U orth.	Np orth.	Pu mon.	Am dhcp	Cm dhcp	Bk dhcp	Cf dhcp	Es	Fm	Md	No	Lr

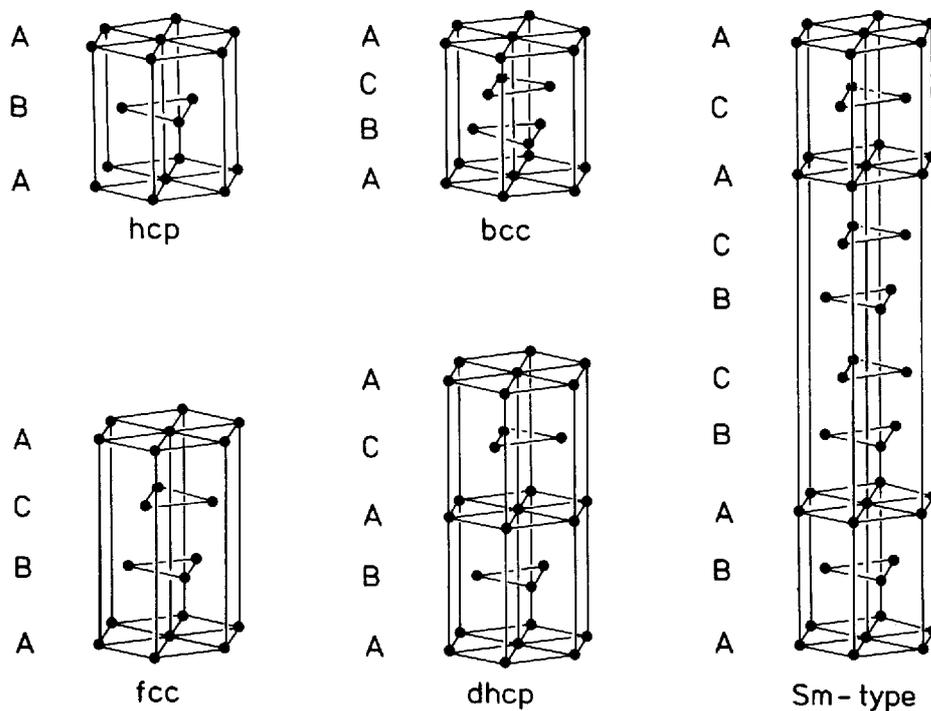
Fig. 1. Crystal structures of the metallic elements at low temperature.

layers of spheres of equal radii in the closest possible fashion. As a result of this close-packing the coordination number in these four structures is 12, each atom being surrounded by 12 nearest neighbours. The bcc structure is a little less close-packed and has a coordination number of 8, although it is sometimes referred to as having a coordination number of 14 on account of the 6 next nearest neighbours, which are only slightly farther away than the nearest neighbours.

It may be seen from Fig. 1 that the crystal structures of the metallic elements tend to occur in sequences when viewed as functions of atomic number or hydrostatic pressure. The most prominent example of this phenomenon occurs with the d transition metals, where all three transition series, excluding the four magnetic 3d metals, exhibit the same hcp→bcc→hcp→fcc sequence as the d states become progressively filled. A similar sequence is found in the lanthanides where the hcp→Sm-type→dhcp→fcc sequence established as a function of decreasing atomic number may also be realized by subjecting each individual lanthanide metal, except Ce, Eu, and Yb, to hydrostatic pressure. Finally, the alkaline earth metals, together with the divalent rare earths Eu and Yb, are part of a short fcc→bcc sequence which is also realized in Ca, Sr, and Yb under high pressure.

In the present contribution we shall establish the extent to which the systematics outlined above can be explained by means of a state-of-the-art theory for the ground state of the bonding electrons. The theory we apply is a one-electron theory in which each electron is treated as an independent particle moving in the effective potential from all the other electrons and the nuclei, and the only input to the calculations is the atomic number of the metal to be treated. In order to be able to reduce the original many-body problem significantly one has to solve the electronic structure problem self-consistently, and to this end we use the Linear Muffin-Tin Orbital (LMTO) method (Andersen 1975) in conjunction with a scaling principle as outlined by Skriver (1984). The structural energy differences which determine the relative stability of the crystal structures to be studied are in turn obtained from the one-electron states by means of Andersen's force theorem (Mackintosh and Andersen 1980). The whole procedure is quite general and allows us to treat all metals on the same footing.

Fig. 2. Close-packed crystal structures of the elemental metals.



The remainder of the present contribution is organized as follows: In Sect. 1.1 we outline the simplest possible theory of structural stability in terms of the density of electronic states, and in the following section 1.2 we apply this simple theory to state densities obtained by means of canonical band theory. In Sect. 2 we review previous theoretical efforts in the field and compare them with the present approach, the theoretical foundations of which are discussed in Sect. 3. In Sect. 4 we outline an electrostatic correction to the Atomic Sphere Approximation (ASA) which becomes important for structures less close-packed than those shown in Fig. 2. Finally, in Sect. 5 we present the calculated structural energy differences for the alkalis, the alkaline earths, the transition metals, the lanthanides, and the light actinides.

1.1. A simple theory of structural stability

In the main part of the following we shall describe the results of a series of calculations of the relative stability of the crystal structures of some 40 elemental metals. In such a presentation, centred around an account of theoretical results and their relation to experimental observations, it is easy to lose track of the principles upon which the calculations are based. We shall therefore immediately present a simple model which will illustrate these principles and in addition will serve to make more comprehensible the complete calculations to be described later.

According to standard textbooks one may imagine a metal formed in the thought experiment illustrated in Fig. 3 where N initially infinitely separated metal atoms are slowly brought together. Here we shall consider specifically a transition metal in which the important states have d character. As a result of the increasing contact between neighbouring atoms the $5N$ atomic d states give rise to a band of energies ranging from B which corresponds to bonding between most neighbours to A which corresponds to antibonding between them. The band of energies formed in this way constitutes the energy band of the metal, and it contains all the one-electron states which the conduction electrons may occupy.

The energy gained in the above process is called the cohesive energy, and according to Fig. 3 it is simply the difference between the total atomic energy nE_a and the total band energy $n\bar{E}$, i.e.

$$E_{\text{coh}} = n(E_a - \bar{E}) \quad (1)$$

assuming an occupation of n d electrons per atom. In writing down (1) we have furthermore assumed that the d states broaden around the atomic level E_a , that is that the centre C of the d band coincides with E_a .

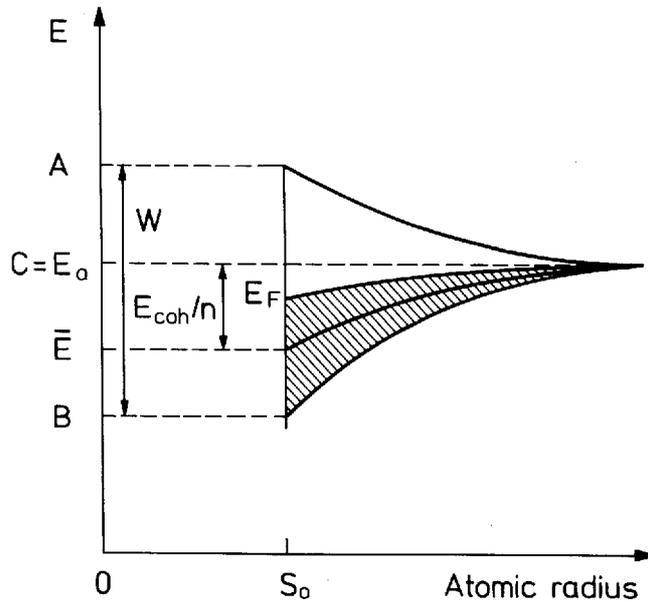


Fig. 3. Formation of the energy band of a metal from an atomic energy E_a . The width is W , the bottom and top B and A , respectively, the Fermi level, i.e. the highest occupied energy E_F , the cohesive energy E_{coh} , and the number of electrons per atom n .

The average energy \bar{E} , which corresponds to the centre of gravity of the occupied part of the d band and which enters (1), may be obtained by summing the one-electron energies ε_i between the bottom of the band B and the highest occupied one-electron level E_F , i.e.

$$\begin{aligned}\bar{E} &= n^{-1} \sum_i^{\text{occ}} \varepsilon_i \\ &= n^{-1} \int^E E N(E) dE\end{aligned}\quad (2)$$

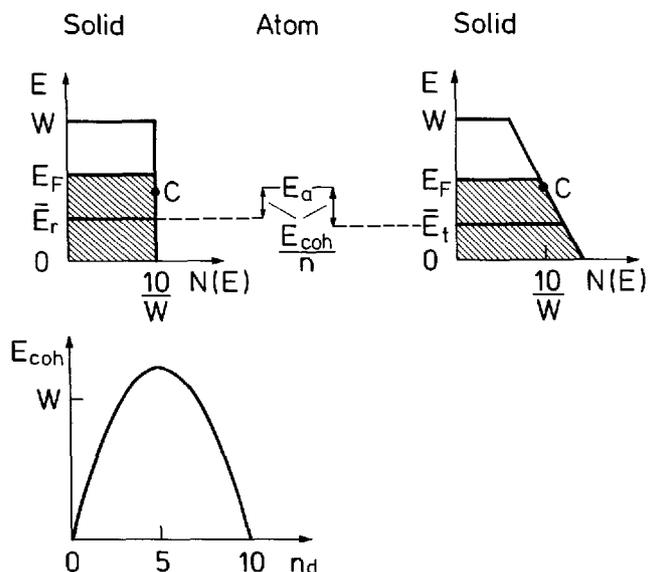
where we have introduced the state density function $N(E)$ which describes how the states are distributed in the energy range from B to A .

If we assume that all states within the d band are equally probable the state density will have the rectangular shape shown in Fig. 4a, and the cohesive energy will simply be given by

$$E_{coh}^r = \frac{W}{20} n(10-n)\quad (3)$$

As noted by Friedel (1969) this form clearly exhibits the parabolic variation with the d occupation, cf. Fig. 4, which is also found experimentally (Gschneidner 1969, Friedel and Sayers 1977), especially when proper account is taken of the atomic effects (Brooks and Johansson 1983), and

Fig. 4. Rectangular and skew state densities modelling the dependence of the cohesive energy upon crystal structure. The cohesive energy as a function of d occupation n is shown for the rectangular state density at the bottom.



this agreement was taken as confirmation of the assumptions of the model outlined above.

From Eq. (1) and Fig. 4 it is clear that the energy gained in forming a metal from the free atoms depends upon the relative position of the atomic d level E_a and the average band energy \bar{E} . The latter depends upon the shape of the state density which in turn depends upon the arrangement of the atoms in the metal crystal, and hence different crystal structures will lead to different cohesive energies.

It follows that the relative stability of all possible crystal structures for a given metal will be determined by the particular shape of the corresponding state densities. We have illustrated this simple result in Fig. 4 from which it is straightforward to see that the crystal structure leading to the skew state density will have a higher cohesive energy and hence be more stable than the structure which leads to the rectangular state density on account of the lowering of \bar{E} . Hence, the relative stability of two crystal structures may be estimated simply by comparing the corresponding average band energies \bar{E} .

In the complete calculations to be reported later we have applied this simple principle to accurately calculated state densities, and the success with which the results explain the experimental observations may be

taken as a justification of the assumptions underlying the one-electron approach outlined above. As will be explained in Sect. 3 there is however also theoretical justification for such a one-electron approach in the form of the so-called force theorem (Mackintosh and Andersen 1980) which dictates how the band structures and the corresponding state densities of the metals in the different crystal structures should be calculated.

1.2. Structural stability from canonical band theory

The concept of canonical bands (Andersen 1975, Andersen and Jepsen 1977) gives rise to a simple and yet realistic procedure for estimating the relative stability of the close-packed crystal structures which form for instance the transition metal sequence, Fig. 1. According to canonical band theory an unhybridized, pure l band may be obtained from (Andersen and Jepsen 1977, Skriver 1984)

$$E_{li}(\mathbf{k}) = C_l + \frac{1}{\mu_l S^2} \frac{\int_{\tilde{S}_l}^{\mathbf{k}}}{1 - \gamma_l \int_{\tilde{S}_l}^{\mathbf{k}}} \quad (4)$$

where $\int_{\tilde{S}_l}^{\mathbf{k}}$ are the canonical bands which depend solely upon the crystal structure, S is the atomic Wigner-Seitz radius, C_l the centre of the l band, μ_l the band mass, and γ_l a distortion parameter. The three potential parameters C_l , μ_l and γ_l depend upon potential and volume but not upon crystal structure.

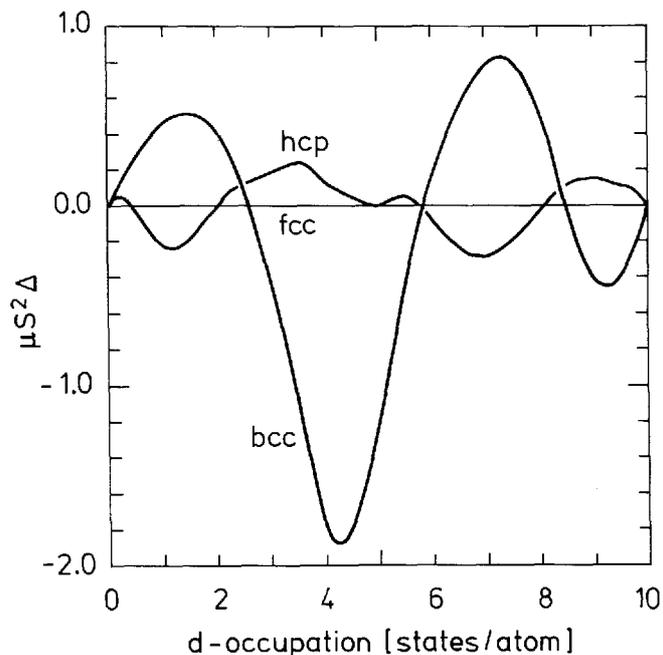
In a transition metal one may to a good approximation neglect all but the d bands. Since furthermore γ_d is small, one has the following potential-, i.e. atomic number-, independent estimate of the band contribution to the cohesive energy E_{coh}

$$\begin{aligned} \mu_d S^2 E_{\text{coh}} &= - \mu_d S^2 \int^{E_F} (E - C_d) N_d(E) dE \\ &= - \int \mathcal{J}_d(n_d) \int_{\tilde{S}_d} \tilde{N}_d(\tilde{S}_d) d\tilde{S}_d \end{aligned} \quad (5)$$

in terms of the first-order moment of the canonical state density \tilde{N}_d . Andersen et al. (1977) have evaluated (5) as a function of d occupation number n_d and found the expected parabolic behaviour (Friedel 1969) which may also be obtained directly if $N_d(E)$ is approximated by a rectangular state density as explained in the introduction.

Since the centre C_d and the band mass μ_d are independent of crystal structure, the first-order moment (5) may be used to estimate the structural energy differences according to Eq. (9). The result shown in Fig. 5 is identical to that of Andersen et al. (1977) and similar to the one

Fig. 5. Structural energy differences obtained from canonical d bands by means of Eq. (5) as functions of the calculated canonical d occupation.



obtained by Pettifor (1970). It accounts qualitatively for the crystal structures of the non-magnetic transition metals, Fig. 1, in the beginning of the series but fails to predict the fcc structure at high d occupations. This failure is attributed either to a failure of the force relation (Mackintosh and Andersen 1980) or to hard-core effects (Pettifor 1970, 1972, 1977) omitted in Eq. (5).

The lanthanide metals are found to have d occupation numbers varying almost linearly with atomic number from 1.99 in La to 1.45 in Lu (Skriver 1983) or from 2.5 to 2.0 if hybridization is neglected (Duthie and Pettifor 1977). Furthermore, their crystal structures are as closely packed as are those of the d transition metals and hence their structural energy differences may be estimated by Eq. (5). The results shown in Fig. 6 are qualitatively similar to but on the average a factor 1.7 smaller than those obtained by Duthie and Pettifor (1977). In this comparison one may take the d -band width to be approximately $25/\mu_d S^2$ in order to bring their Fig. 2 onto the scale of Fig. 6. The results in Fig. 6 account qualitatively for the hcp→Sm-type→dhcp sequence found experimentally in going from Lu to La and more importantly perhaps, since the d occupation for the lanthanides is calculated to increase with pressure and

decrease with atomic number, they also explain that part of the same sequence is realized when a particular lanthanide metal is subjected to pressure. It therefore follows that the d occupation number, which is essentially a measure of the relative position of the s and d bands, may be used to rationalize the structure of the generalized phase diagram for the lanthanides constructed by Johansson and Rosengren (1975).

At the present stage one should realize that the results obtained by canonical band theory and shown in Figs. 5 and 6 are only qualitative. Indeed, if one considers Fig. 7 where the canonical estimates are compared with experimental crystal structures, one finds that the canonical theory in several cases does not predict the correct crystal structure independently of whether one uses the self-consistent d occupation num-

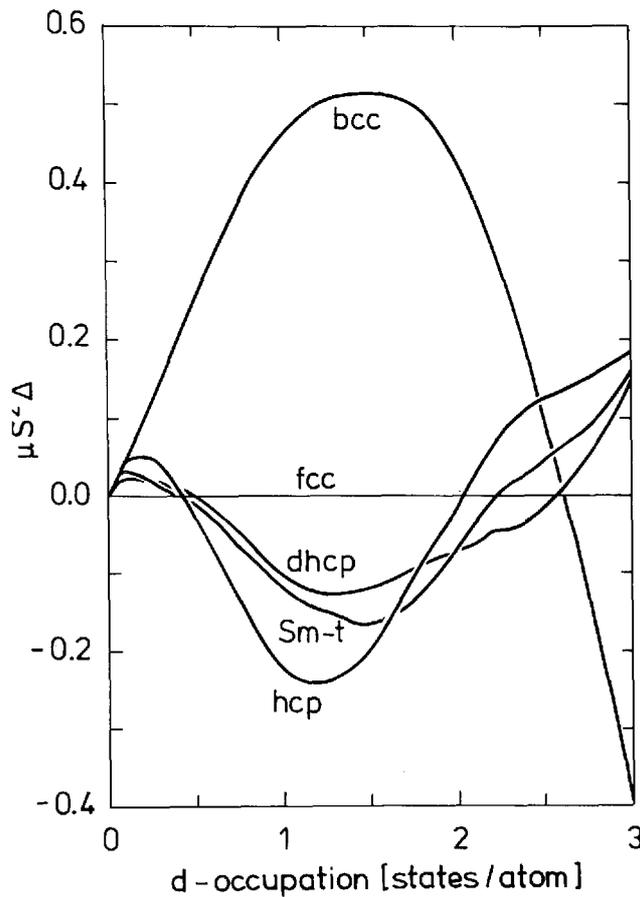


Fig. 6. Structural energy differences obtained from canonical d bands by means of Eq. (5) in the d occupation number range appropriate to the lanthanide crystal structure sequence.

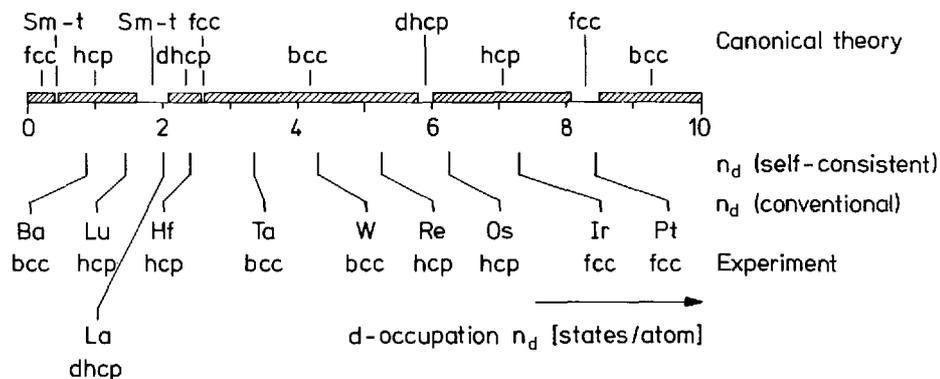


Fig. 7. Canonical estimate of the most stable close-packed crystal structure as a function of the calculated d occupation number compiled from Figs. 2 & 3, horizontal bars. Below are given two estimates of the actual d occupation numbers of the 5d metals together with the experimentally observed crystal structures.

bers or those obtained conventionally by nonlinear interpolation along a row in the periodic table (see Fig. 1). La, Re, and Ir, for instance, are examples of incorrect predictions, but here one may argue that the correct crystal structure is nearby and hence the failure of the theory may be considered less important. Ba is another example and in this case there is no nearby bcc structure. However, in Ba the d occupation number is only a fraction of the total number of electrons and hence a theory based solely upon unhybridized d bands is probably inapplicable. The most important failure is connected with the d occupation range from 1.6 to 2.6 [states/atom]. According to Fig. 7, La, Pr, Nd, and Pm should incorrectly form in the Sm-type structure while Ti, Zr, and Hf are expected to be part of the lanthanide sequence. Instead, the latter three metals form in the hcp structure which is the least stable among those considered in the d occupation range above 2 [states/atom].

It may be concluded that the simple estimate of structural energy differences obtained by means of the first-order moments of the canonical state densities is of limited value as a predictive tool. It is, however, of sufficient physical significance to warrant a study of the crystal structures of metals using a more accurate one-electron theory, and to be used in the interpretation of the results of such a study.

2. Theoretical approaches to structural stability

The most prominent crystal structure sequence in the periodic table is the hcp→bcc→hcp→fcc sequence found among the d transition metals, see Fig. 1. Qualitative explanations of this trend have been given by Brewer (1967) in terms of Engel correlations between the valence sp electrons and by Kaufman and Bernstein (1970) in terms of semi-empirical thermodynamic calculations of phase diagrams, whereas Deegan (1968), Dalton and Deegan (1969), and Ducastelle and Cyrot-Lackmann (1971) have attempted more quantitative approaches based upon one-electron theory.

Deegan (1968) and Dalton and Deegan (1969) showed that the stability of the bcc phase for nearly half-filled d shells might be explained by differences in the sum of one-electron band structure energies, and they pointed to the special double-peak structure of a bcc state density as responsible for this stability. Later, Pettifor (1970, 1972) extended the work of Dalton and Deegan and showed that the entire crystal structure sequence of the transition metals could be accounted for by a one-electron approach. In his calculations Pettifor (1977) found no evidence for the Brewer-Engel correlation (Brewer 1967), which relates crystal structure stability to the sp occupation numbers, and instead he related the hcp→bcc→hcp→fcc sequence to the change in d occupation which takes place across a transition series. This latter viewpoint has proven to be very fruitful in that it may be used as a simple »one-parameter theory« which in many cases provides remarkably good estimates of structural stabilities also for non-transition metals such as the alkaline earths (Skriver 1982).

The crystal structures of the trivalent lanthanides, i.e. Pr through Lu except Eu and Yb, exhibit such regularity, as functions of atomic number, pressure, and temperature, that Johansson and Rosengren (1975) were able to construct a single generalized phase diagram for these metals. In this case the crystal structures observed under ambient conditions, (see e.g., Beaudry and Gschneidner 1978) are found to be part of the sequence hcp→Sm-type→dhcp→fcc→fcc' established by high-pressure experiments (Jayaraman and Sherwood 1964, Piermarini and Weir 1964, Jayaraman 1965, McWhan and Stevens 1965, 1967, Liu et al. 1973, Liu 1975, Nakae 1978) and alloying (Koch 1970). Here fcc' refers to the recently discovered distorted fcc structure (Grosshans et al. 1981). The lanthanide sequence is also found in Y (Vohra et al. 1981) where there are no occupied f states and in the heavier actinide (Stephens et al. 1968,

Akella et al. 1979, 1980, Roof et al. 1980, Roof 1982, Benedict et al. 1984) at pressures where the 5f states are still localized. Qualitative explanations of the hcp→Sm-type→dhcp→fcc sequence have been attempted in terms of pseudopotential theory by Hodges (1967) and in terms of a 4f contribution to the bonding by Gschneidner and Valletta (1968), while Duthie and Pettifor (1977) gave a quantitative explanation in terms of one-electron theory.

Duthie and Pettifor (1977) showed that the lanthanide crystal structure sequence could be explained by differences in the total one-electron band structure energies, and they found a strong correlation between crystal structure and d-occupation number. Hence it appears that the lanthanide metals, as far as their crystal structures are concerned, behave as ordinary 5d transition metals with a d occupation ranging from approximately 2.0 in La to 1.5 in Lu. This result is very appealing because there is a one-to-one correspondence between the calculated d-occupation number and the single f parameter used by Johansson and Rosengren (1975) and Johansson (1978) to rationalize the lanthanide crystal structure sequence, and because it is immediately possible to understand the behaviour of Y (Vohra et al. 1981) and the heavy actinides (Stephens et al. 1968, Akella et al. 1979, 1980, Roof et al. 1980, Roof 1982, Benedict et al. 1984) within the same framework.

At first sight it may seem surprising that the crystal structures of so many metals can be explained on the basis of differences in the total one-electron band structure energies alone, since the total electronic energy, apart from the one-electron term, has contributions also from double counting and exchange-correlation. However, it has recently been shown (Andersen et al., 1979, Mackintosh and Andersen 1980, see also page 119 of Heine 1980) that, provided the one-electron *potential* is kept frozen upon a displacement of the atoms, the corresponding changes in the double counting and exchange-correlation terms cancel to first order in the appropriate local electron density, and hence the difference in the sum of the one-electron energies, obtained by means of the frozen, i.e. not self-consistently relaxed, potential, will give an accurate estimate of the corresponding self-consistent change in the total electronic energy. It is exactly this cancellation, which also leads to the so-called pressure expression (Nieminen and Hodges 1976, Pettifor 1976) and to the more general force relation derived by Andersen (see Mackintosh and Andersen 1980), that in turn justifies the simple band structure approach taken for instance by Pettifor (1970, 1972, 1977).

In their work Pettifor (1970, 1972, 1977) and Duthie and Pettifor

(1977) focused their attention on the contribution to the total energy from the d bands and either neglected hybridization with the sp bands entirely or included hybridization appropriate to some average element. Hence their picture is essentially a canonical one (cf. Sect. 1.2) in which the energy band structures depend only upon crystal structure and not upon band-filling. It is obvious that such a picture, although adequate for the d transition metals, will fail in cases where states of non d character are as or more important than the d states, as they are for instance in the alkali, the alkaline earth and light actinide metals. Fortunately, the force theorem is not restricted to the canonical approximation and it has recently been used in theoretical investigations of crystal structures in the third row metals (Moriarty and McMahan 1982, McMahan and Moriarty 1983) the alkaline earth metals (Skriver 1982), and in Cs above the s-d transition (McMahan 1984).

In the present work we go beyond the canonical approximation and use the force theorem (Mackintosh and Andersen 1980) to calculate the structural energy differences for all the 3d, 4d, and 5d transition metals at zero pressure and temperature. In addition we investigate the effect of hydrostatic pressure upon the crystal structures of alkali, alkaline earth, lanthanide and actinide metals.

Traditionally the non-transition metals, e.g. alkali and alkaline earth metals, have been treated by means of pseudopotential theory, and the crystal structures predicted from this approach are generally in good agreement with experiment (Animalu 1967, Heine and Weaire 1970, Moriarty 1973, 1982, Hafner and Heine 1983, Young and Ross 1984). It has, however, not been straightforward to generalize the pseudopotential method to treat narrow d band materials, and to do so one has had to add localized orbitals to the plane-wave basis (Zunger and Cohen 1979). Thus the d band in K is described by the d component of plane-waves while that of Cu is described by additional d orbitals, which is somewhat inconsistent with the smooth lowering of the 3d band through the series K, Ca, Sc, ..., Cu. The method has, however, proved to be very accurate.

The present approach, based upon the Linear Muffin-Tin Orbital (LMTO) method (Andersen 1975), has the advantage of employing the same type of basis functions for all the elements thus leading to a conceptually consistent description of trends throughout the periodic table. In addition, the LMTO method is extremely efficient on a computer requiring only the solution of an eigenvalue problem of 9×9 (or 16×16 if f states are included) per atom at each point in reciprocal space. Since we

are mainly interested in trends, we have neglected the nonspherical contributions to the charge density, which may explain what seems to be a systematic overestimate of the calculated structural energy differences. We have furthermore neglected a structure-dependent electrostatic interaction between atomic spheres except in the few cases where it contributes significantly to the energy differences.

3. One-electron theory of structural stability

At low temperatures the crystal structure of a metal is determined by the total electronic energy U in addition to a small contribution from the zero-point motion^{*}, which we shall neglect. Hence, if one wants to determine the stability of some crystal structure, say bcc, against some reference structure, which we shall take to be the close-packed fcc structure, one may calculate the total energy of both phases and form the structural energy difference

$$\Delta_{\text{bcc-fcc}} = U_{\text{bcc}} - U_{\text{fcc}} \quad (6)$$

where the total energy according to the local density approximation (Kohn and Sham 1965) may be written as the sum over occupied states of the one-electron energies ϵ_i corrected for double counting, plus electrostatic terms (see e.g., sections 13 and 15 of Heine 1980 or sections 7.2 and 7.3 of Skriver 1984), i.e.

$$U = \sum_i^{\text{occ}} \epsilon_i - \text{double counting} + \text{electrostatic} \quad (7)$$

If the difference (6) is negative the bcc structure will be stable against fcc.

The total energy for say a 4d metal is of the order of 10^4 [Ry] mainly because of the contributions from the low-lying core levels while typical structural energy differences are of the order of 10^{-3} [Ry]. Hence, extreme accuracy is needed in order to use (6) directly, and one would like to have a numerically more satisfactory procedure. The force theorem (Mackintosh and Andersen 1980) gives rise to such a procedure, but more importantly perhaps it casts the problem of finding stable crystal

^{*} The zero-point energy is proportional to the Debye temperature i.e. $E_0 = (9/8) k_B \Theta_D$. Typically Θ_D varies by 1-10 [K] between different structures of the same metal (see Gschneidner 1964) and hence the ΔE_0 to be added to (6) is of the order of 0.01-0.1 [mRy] which in most cases will be too small to affect the structural stabilities.

structures into a form where the significant contribution comes from the one-electron valence energies and not from double counting nor from the deep core-levels.

The force theorem dictates that we adopt the following procedure: For a given metal at a given atomic volume one must solve the energy-band problem self-consistently assuming the reference crystal structure. To this end we use the LMTO method (Andersen 1975) within the Atomic Sphere Approximation (ASA) including the combined correction to the ASA (Andersen 1975). In addition we take account of the relativistic effects, except spin-orbit coupling which we neglect, include exchange-correlation in the form given by von Barth and Hedin (1972), and freeze the appropriate cores. This part of the calculations is described in detail by Skriver (1984). We have now minimized the energy functional $U\{n\}$ with respect to changes in the electron density n and obtained the ground state density n_{fcc}^{sc} . Because of the stationary properties of U one may obtain, for instance, U_{bcc} from a trial charge-density n_{bcc}^{tr} constructed by positioning the self-consistent fcc atomic-sphere *potentials* in a bcc geometry, solving the one-electron Schrödinger equation, and populating the lowest-lying one-electron states. Hence,

$$\Delta_{bcc-fcc} = U_{bcc}\{n_{bcc}^{tr}\} - U_{fcc}\{n_{fcc}^{sc}\} \quad (8)$$

where the errors relative to (6) are of second order in $n_{bcc}^{tr} - n_{bcc}^{sc}$. Now, the use of a *frozen*, i.e. not self-consistently relaxed, *potential* to generate n_{bcc}^{tr} ensures that the chemical shifts in the core levels drop out of Eq. (8) and also that the double-counting terms cancel. Hence, the core level energies and the double-counting terms may be neglected entirely in Eq. (7) leaving only the valence one-electron energies and the electrostatic terms to be considered. The fact that the freezing of the potential leads to such a computationally simple and conceptually important result was already noted by Pettifor (1976) in his derivation of the pressure expression.

Within the atomic sphere approximation (Andersen 1975) the atomic Wigner-Seitz sphere of an elemental metal is neutral and there is therefore no electrostatic interaction between the spheres. Hence the electrostatic terms in Eq. (7) vanish and the structural energy difference (8) may be obtained from

$$\Delta_{bcc-fcc} = \int^{E_F} E N_{bcc}(E) dE - \int^{E_F} E N_{fcc}(E) dE \quad (9B)$$

where $N(E)$ is the one-electron state density. Furthermore, the ASA allows a separation of the potential- and crystal-structure-dependent

parts of the energy band problem (Andersen 1975, Andersen and Jepsen 1977, Skriver 1984). Hence, all that is required at a given atomic volume, in addition to the self-consistent fcc calculation, is to calculate the energy bands of the relevant crystal structures with the use of the self-consistent fcc potential parameters, evaluate the sums of the one-electron energies, and subtract according to Eq. (9). This procedure is quite general, treats all s, p, d, and f electrons on the same footing, and may be applied to all metals in the periodic table.

4. Madelung correction to the ASA

The errors of neglecting the structure-dependent electrostatic terms in (7) may be estimated by means of what has been called either the Muffin-Tin (Glötzel and Andersen, unpublished) or Ewald (Esposito et al. 1980) correction to the ASA. To derive this correction one observes that the electrostatic energy per ion of a lattice of point ions of charge $q_s|e|$ embedded in a negative neutralizing uniform charge density is given by the well-known Madelung expression

$$U_M = - 1/2(q_s|e|)^2 \frac{\alpha_M}{S} \quad (10)$$

where α_M is the lattice Madelung constant and S the atomic Wigner-Seitz radius. In the ASA this expression is approximated by the energy of an ion embedded in a single neutralizing atomic sphere, whereby $\alpha_M(\text{ASA}) = 1.8$. The correction is therefore

$$\Delta U_M = 1/2(q_s|e|)^2 \frac{1.8 - \alpha_M}{S} \quad (11)$$

In a Muffin-Tin model the effective charge $q_s|e|$ is the charge density in the interstitial region between the Muffin-Tin spheres times the volume of the unit cell. In the ASA this becomes

Table 1. Madelung constant to be used in Eq. (11).

	α_M	$1.8 - \alpha_M$	$(1.8 - \alpha_M) - (\)_{fcc}$
fcc	1.79174723	8.253 10^{-3}	
bcc	1.79185851	8.142 10^{-3}	- 0.111 10^{-3}
hcp	1.79167624	8.324 10^{-3}	0.071 10^{-3}
$\alpha\text{-U}^*$)	1.78418298	15.817 10^{-3}	7.564 10^{-3}

*) $b/a = 1.964$, $c/a = 1.709$, $y = 0.1$

$$q_s|e| = \frac{4\pi}{3} S^3 n(S) |e| \quad (12)$$

where $n(S)$ is the electron density at the atomic radius.

For close-packed crystal structures α_M is approximately 1.8, see Table 1, and hence the correction (11) is smallest in these. Typically q_s^2/S lies in the range from 0.5 to 5 [a.u.] so that the Madelung correction for the bcc and hcp structures relative to the fcc structure lies in the range 0.05 to 0.5 [mRy].

5. Structural stability from LMTO band calculations

In the following we shall present structural energy differences for most metallic elements to the left of and including the noble metals as obtained by means of the procedure described in Sect. 3. The results will be valid only at low temperature and at atmospheric pressure, strictly $T = 0$ [K] and $P = 0$ [GPa], except in a few important cases where structural stability has been followed as a function of pressure.

5.1. The alkali metals

The calculated structural energy differences for alkali metals at equilibrium are almost two orders of magnitude smaller than those of, for instance, the alkaline earth metals. To judge the accuracy of our approach we have therefore studied these differences as functions of pressure as shown in Fig. 8 from equilibrium down to a compression of 2.5. The results in Fig. 8 include the Madelung correction (11) which turns out to be crucial in the comparison with recent pseudopotential and LMTO results (Moriarty 1982, Moriarty and McMahan 1982, McMahan and Moriarty 1983).

From Fig. 8 it is expected that the heavy alkalis at low temperature and pressure should form in the bcc structure while Li should be hcp. Experimentally it is known (Donohue 1975, Young 1975) that all five alkali metals at room temperature form in the bcc structure, and that they remain in this structure down to 5 [K] except Na which below 51 [K] transforms into the hcp structure and Li which at low temperature exhibits both an hcp and an fcc phase. Hence, except for Na the low pressure structures are correctly predicted.

Recently, Moriarty (1982) successfully estimated the structural stability for some 20 non-transition metals by means of his Generalized Pseudopotential Theory (GPT). He found incorrectly (see his Table VIII)

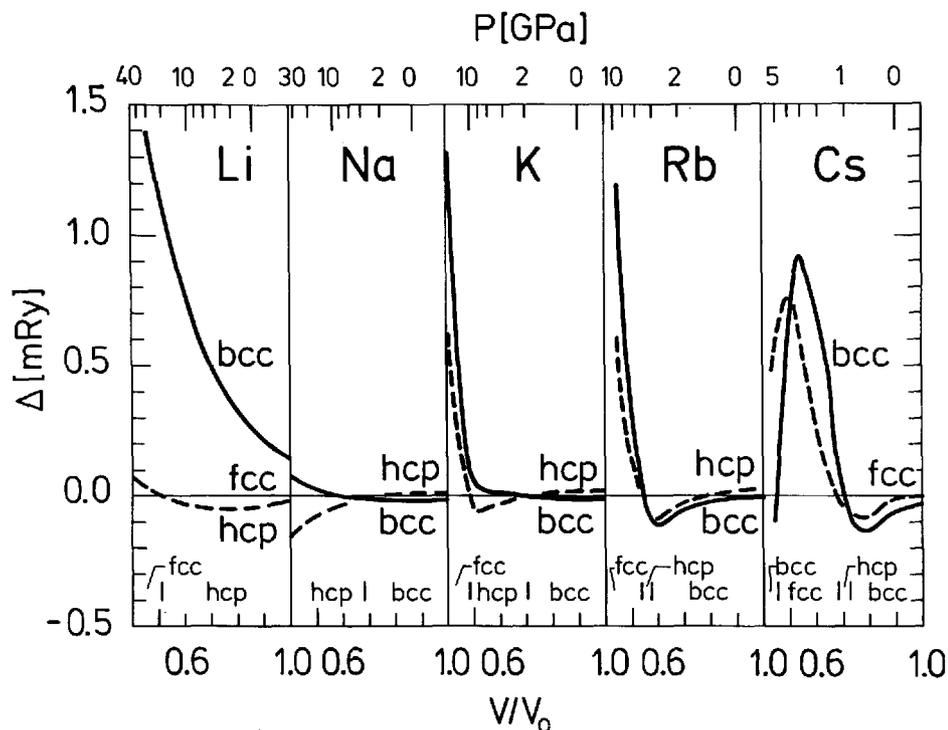


Fig. 8. Structural energy differences for the 5 alkali metals as functions of the relative volume V/V_0 . At the top is given the calculated LMTO pressure P . The calculations included s , p , and d orbitals and the Madelung correction Eq. (11).

that all the alkali metals at $P = 0$ and $T = 0$ should form in the hcp structure, but pointed out that at a slight compression the experimentally observed bcc structure would be stable in the heavy alkalis K, Rb, and Cs. A similar problem is encountered in another recent pseudopotential study (Young and Ross 1984) where the structures of Li and K at low temperature and pressure are predicted in agreement with experiment but where Na is expected to be fcc. On the other hand, in view of the extremely small energies involved, see Fig. 8, it is not surprising that the prediction of the low-pressure part of the alkali phase diagrams is a severe test of any calculation.

In their work on the third-row metals McMahan and Moriarty (1983) compared structural energy differences obtained by means of the LMTO and GPT methods and found excellent qualitative agreement except for Na. If we compare our Na results in Fig. 8 with their Fig. 1 we find

somewhat surprisingly that our calculations are in closer agreement with their GPT than with their LMTO results. There are several reasons for the differences between the two LMTO calculations. Firstly, we have included the Madelung correction (11) without which the calculated bcc curve is entirely above and the hcp curve entirely below the fcc, in qualitative agreement with their LMTO results. Secondly, we have sampled the Brillouin zone on a finer mesh, i.e. 916, 819, and 448 points in the irreducible wedge for fcc, bcc, and hcp, respectively, and finally, we have improved the convergence of the reciprocal lattice sums in the expression for the combined correction terms (Andersen 1975) whereby the numerical errors in the structural energy differences for Na are below 0.01 mRy. As a result it appears that in the case of closely packed crystal structures the LMTO method including the Madelung correction (11) has an accuracy comparable to that attained by pseudopotential theory.

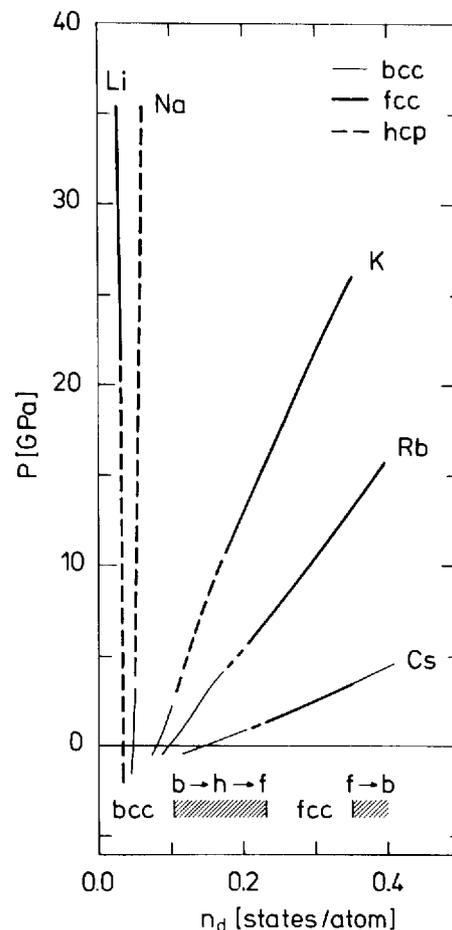
Owing to the inclusion of only three crystal structures in Fig. 8, Cs is incorrectly calculated to transform into the bcc structure at a compression of 2.2. However, in a recent study of Cs above the s-d transition, i.e. beyond the pressure range of the present work, McMahan (1984) found that Cs had transformed into the Cs IV structure before the bcc structure became more stable than fcc, in agreement with high pressure experiments (Takemura et al. 1981, 1982).

5.2. *The alkali metals at moderate compression*

According to Fig. 8 all the alkali metals should at low temperature be part of the same crystal structure sequence $\text{bcc} \rightarrow \text{hcp} \rightarrow \text{fcc}$, and one would anticipate that these transitions are driven by the pressure-induced lowering of initially unoccupied d states through the Fermi level, whereby electrons are gradually transferred from the s into the d band. If one plots the calculated crystal structures as functions of d occupation number as in Fig. 9 it is seen that only in the heavy alkalis K, Rb, and Cs is this mechanism at work while the transitions in Li and Na at least below 35 [GPa] have a different origin.

The experimental situation at room temperature has recently been summarized as follows (Takemura and Syassen 1983, Olijnyk and Holzapfel 1983). Li exhibits a $\text{bcc} \rightarrow \text{fcc}$ transition at 6.9 [GPa] (Olinger and Shaner 1983) while Na remains in the bcc structure up to at least 30 [GPa] (Alexandrov et al. 1982) which substantiates the notion that the s-d transition is unimportant in these two metals. The heavy alkalis all exhibit a $\text{bcc} \rightarrow \text{fcc}$ transition [K (Takemura and Syassen 1983, Olijnyk and Holzapfel 1983), Rb (Takemura and Syassen 1982), Cs (Hall et al.

Fig. 9. Calculated crystal structures for the alkali metals as functions of the LMTO pressure and d occupation number.



1964)] before they transform into more complex structures of which only the so-called Cs IV has been solved so far (Takemura et al. 1982).

To our knowledge there are no low-temperature high-pressure experiments which could substantiate the existence of the predicted $\text{bcc} \rightarrow \text{hcp} \rightarrow \text{fcc}$ sequence, where according to Figs. 8 and 9 the hcp phase at least in K should be stable over an appreciable pressure range. However, in view of the fact that temperature at atmospheric pressure stabilizes the bcc phase to the extent that all the alkali metals are bcc above 100 K it is reasonable to assume that the intermediate hcp phase, which is only marginally stable, is also suppressed at higher temperatures. Thus, in a high-pressure experiment at room temperature one would see a direct $\text{bcc} \rightarrow \text{fcc}$ transition, as indeed one has observed (Hall et al. 1964, Take-

mura and Syassen 1982, 1983, Olijnyk and Holzapfel 1983). If the hcp phase is suppressed the best estimate of the room temperature bcc→fcc transition pressure is the critical pressure for the low-temperature hcp→fcc transition (cf. Fig. 8). We find the transition pressures to be 11, 5.5, and 1.4 [GPa] for K, Rb, and Cs, respectively, which should be compared to the experimental values of 11, 7, and 2.2 [GPa] listed in the references cited above.

Independent of whether the intermediate hcp phase exists or not, the high-pressure fcc phase in K, Rb, and Cs is much more stable than the initial bcc phase, see Fig. 8. Bardeen (1938) suggested already in 1938 that the transition observed at 2 [GPa] in Cs was from the normal bcc to an fcc phase and that it resulted from the non-electrostatic interaction energy of the ions, the important term being the Born-Mayer (Born and Mayer 1932) repulsion between the ion cores. Here we shall show that the fcc phase in the heavy alkalis owes its stability directly to the pressure-induced s-d transition which is also shown to be behind, for instance, the isostructural fcc-fcc transition in Cs (Glötzel and McMahan 1979).

In Fig. 10 we compare the important parts of the fcc and bcc band-structures of Cs at the zero-pressure volume, V_0 , and at the volume where the fcc phase becomes more stable than the initial bcc phase. The four band structures may be characterized as nearly free-electron and s-like below the Fermi level E_F and d-like above E_F . Typical d states have symmetry labels such as Γ_{12} , $\Gamma_{25'}$, H_{12} , and X_3 , and they are seen to approach the Fermi level under compression. At $V = V_0$ the fcc and bcc band-structures are found to be extremely similar in the range below E_F which is important in the sums over occupied states in Eq. (9): They are both parabola shaped and »touch« E_F at a single symmetry point, L_1 for fcc and N_1 for bcc. As a result, the sum of the one-electron band-structure energies are almost equal and the main contribution to the stability of the bcc phase comes from the electrostatic Madelung term (11) which is negative, see Table 1.

At $V = 0.7 V_0$ hybridization with the descending d band has moved the X_1 and neighbouring levels below E_F thereby lowering the energy in the fcc phase with respect to that in the bcc phase to the extent that the Madelung term is overcome and the structural energy difference is zero. Under further compression the X_1 level continues to descend and the fcc phase becomes increasingly stable, see Fig. 11. This trend is eventually broken because the maximum in the $\Gamma_1\Delta_1X_1$ band moves away from X and because the X_3 level drops below the Fermi level. Both effects de-

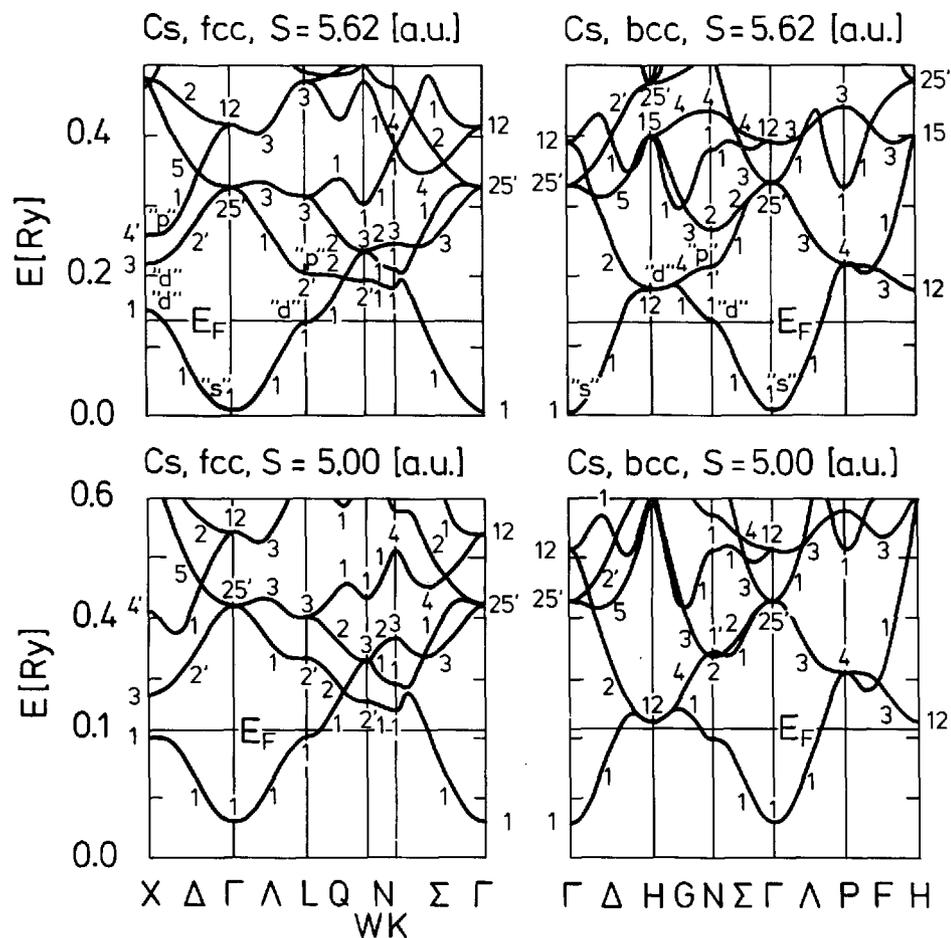


Fig. 10. Energy band structures for Cs at equilibrium, $S = 5.62$ [a.u.], and at a compressed volume, $S = 5.0$ [a.u.]. Conventional symmetry labels are given and the dominant s , p or d character is indicated at a few selected energy levels.

stabilize the fcc structure and subsequently Cs transforms into the Cs IV phase. We shall not discuss this development here but refer to the experimental work of Takemura et al. (1982) and the theoretical treatment of McMahan (1984).

The presence of a gap at X (see Fig. 10) near the Fermi level in the compressed fcc phase which has no counterpart in bcc phase (nor in the hcp phase) stabilizes the fcc phase over the bcc in exactly the manner discussed by Jones in his classical work on the phase boundaries in binary

alloys (Mott and Jones 1936, Jones 1937). The electron states below the gap have their one-electron band-energies lowered and are more densely populated than their free-electron or, here, bcc counterparts. The way the fcc phase is stabilized in Cs under pressure is shown in Fig. 11 where one notes that the stabilization occurs gradually from the point where the X_1 level crosses E_F . Hence, although the fcc phase eventually becomes more stable than the bcc phase because of the presence of the band gap at X, there is no direct relation between the volume ($V = 0.70 V_0$) where the phase transition occurs and the volume ($V = 0.82 V_0$) where the van Hove singularity connected with the X_1 level moves through the Fermi level. This delayed action is characteristic of many electronically driven transitions.

In the discussion of the stability of the fcc phase we have considered only Cs for simplicity, but examination of the band structures for K and Rb shows that the above picture applies equally well to these two metals although there are quantitative differences between K, Rb, and Cs caused by the fact that the zero-pressure position of the initially unoc-

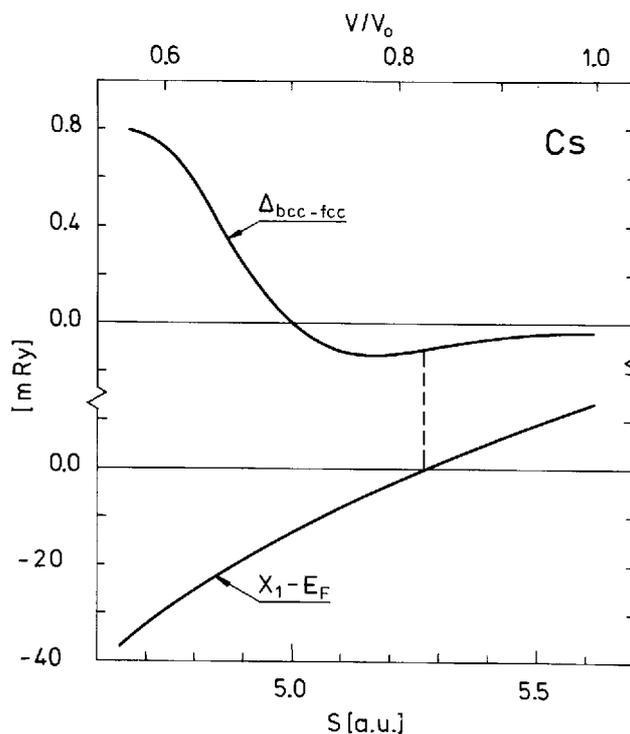


Fig. 11. Structural energy difference $\Delta_{bcc-fcc}$ for Cs, upper panel, and the position relative to the Fermi level, E_F , of the bottom of the gap at X in the fcc structure, lower panel, as functions of atomic radius, S , or relative volume V/V_0 . $V = (4\pi/3)S^3$.

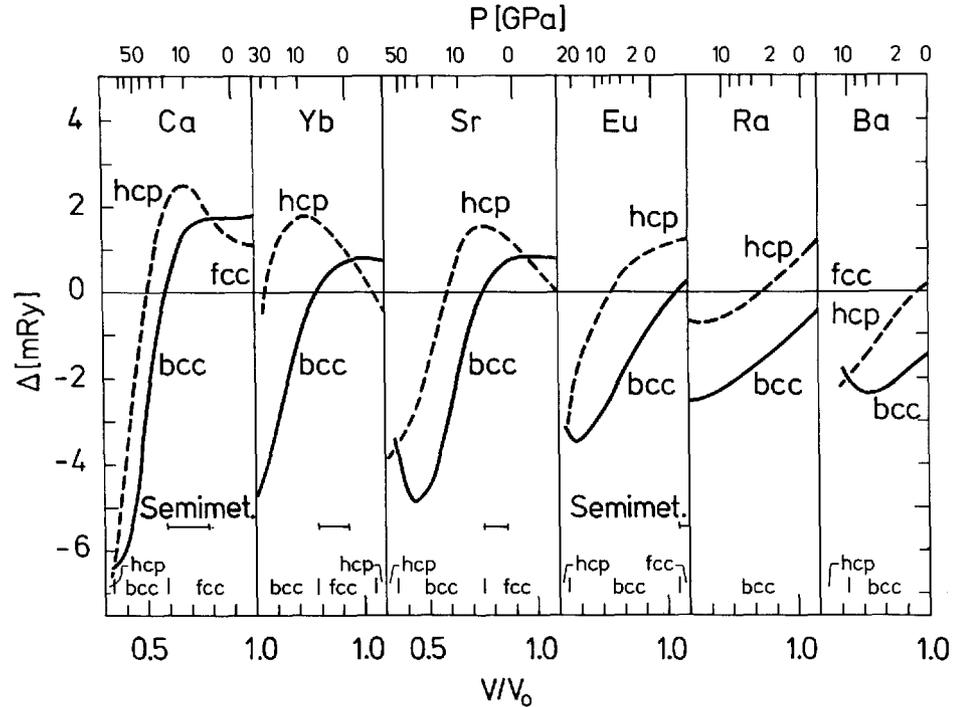


Fig. 12. Structural energy differences for the alkaline earth metals and the two divalent rare earths Eu and Yb as functions of relative volume V/V_0 and LMT0 pressure P . The volume range over which the elements are calculated to be semimetallic is indicated by horizontal bars. The calculations included s , p , and d orbitals but not the Madelung correction Eq. (11).

cupied d band drops relative to the Fermi level as the atomic number increases.

5.3. The alkaline earth metals

The calculated structural-energy differences for the alkaline earth metals under pressure are shown in Fig. 12. In the figure the metals are ordered according to their calculated d occupation number at equilibrium and we have included the two divalent rare earths Eu and Yb, but excluded the divalent metals Be and Mg since they do not really belong to the crystal structure sequence we shall presently be discussing. The results at zero pressure for Be and Mg may, however, be found in the preliminary account (Skriver 1982) of the present work.

According to Fig. 12, Ca, Yb, and Sr at low temperature and pressure should form in the fcc structure while Eu, Ra, and Ba should be bcc.

These predictions are in agreement with experiments (Donohue 1975, Young 1975) except for Yb which at low temperature takes up the hcp structure (Bucher et al. 1970). However, at a slightly expanded volume the hcp phase is calculated to be the stable phase, and hence one may not have to appeal to zero-point motion to explain the anomalous low-temperature hcp phase in Yb. Previous pseudopotential calculations (Animalu 1967) have explained the bcc structure in Ba and the pressure- (and temperature-) induced fcc→bcc transition in Sr, but gave an incorrect (bcc) zero-pressure crystal structure in Ca. Later pseudopotential results (Moriarty 1973) indicated that the stable structure at ordinary pressure should be the fcc structure for all the alkaline earths. Hence, it is still a challenge to pseudopotential theory to predict the crystal structures of the alkaline earth metals as a function of both atomic number and pressure.

There is a strong correlation between the calculated d occupation number and the calculated crystal structure as may be seen in Fig. 13. According to this the heavy alkaline earth metals should be part of the same hcp→fcc→bcc→hcp sequence. At zero pressure each individual metal

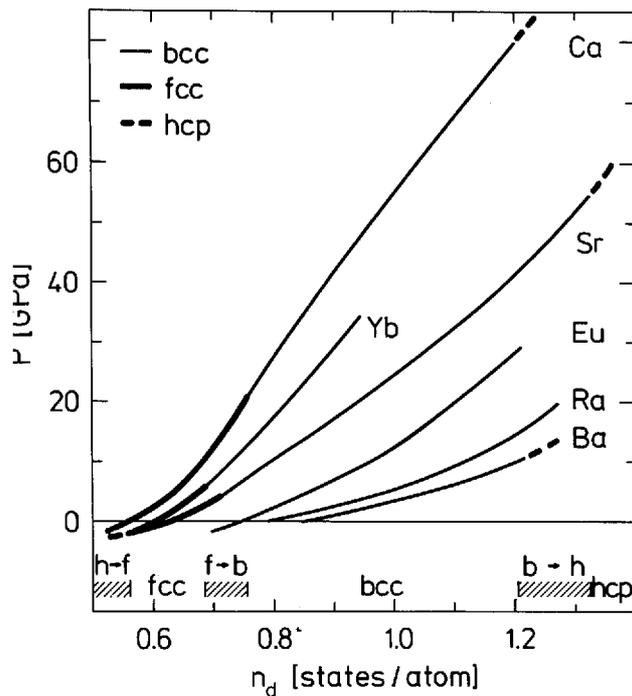


Fig. 13. Calculated crystal structures for the alkaline earth metals as functions of the LMTO pressure and d occupation number.

may be characterized as being at different stages on the continuous s-to-d transition, i.e. by their d occupation number, and the structural phase transitions are then driven by the pressure-induced lowering of the d band with respect to the s band. The correlation is, however, not perfect and the calculated crystal-structure changes occur over a narrow range of d occupation numbers.

Experimentally (Jayaraman et al. 1963a, b, Jayaraman 1964, Olijnyk and Holzapfel 1984) one observes at room temperature the fcc→bcc part of the above sequence but the bcc→hcp transition is found only in Ba whereas the lighter alkaline earth metals transform into more complex high-pressure phases (Olijnyk and Holzapfel 1984) not considered here. The critical pressures for the fcc→bcc transition in Ca, Sr, and Yb plus the bcc→hcp transition in Ba are calculated to be 21, 3.8, 5.5, and 10 [GPa], respectively (cf. Fig. 13). At room temperature Olijnyk and Holzapfel (1984) find experimentally 19.7 [GPa] for the transition in Ca while a low-temperature extrapolation of the high pressure crystallographic measurements by Jayaraman et al. (1983a, b) and Jayaraman (1964) gives 4, 5, and 5 [GPa] for the latter three transitions. In view of the fact that no adjustable parameters have been used to construct Fig. 13, the agreement with the calculated critical pressures may be considered satisfactory.

The band structure calculations show in agreement with the high-pressure resistivity data (Stager and Drickamer 1963a, b, Souers and Jura 1963, McWhan et al. 1963) that Ca, Sr, and Yb in the fcc phase should undergo a metal-semimetal-metal transition under pressure as described in detail for Ca by, for instance, Jan and Skriver (1981). Recently, Dunn and Bundy (1981) re-measured Ca and found the pressure range of the semimetallic phase to be much narrower than that found in earlier measurements (Stager and Drickamer 1963a) or predicted by band theory (McCaffrey et al. 1973, Mickish et al. 1974, Jan and Skriver 1981). Jan and Skriver (1981), for instance, predicted that fcc Ca should be semimetallic from 4 to 29 [GPa]. In the present extension of those calculations it is seen in Fig. 12 that before Ca reaches 29 [GPa] it is expected to transform into the bcc phase whereby the semimetallic behaviour will be terminated already at 21 [GPa]. This termination of the semimetallic phase at approximately 20 [GPa] is in agreement with both resistivity (Dunn and Bundy 1981) and crystallographic (Olijnyk and Holzapfel 1984) measurements. However, the critical pressure of 4 [GPa] for the onset of the semimetallic behaviour is still too low compared to that obtained from the resistivity data of Dunn and Bundy (1981), and this

discrepancy must be due to a failure of local-density theory of the kind mentioned by Jan og Skriver (1981).

In recent high-pressure measurements (Holzapfel et al. 1979, Takemura and Syassen 1985) both Eu and Yb are found to transform from the bcc to the hcp phase in seeming agreement with the systematics exhibited in Fig. 13. However, since Yb (Syassen et al. 1982) and presumably also Eu (Johansson and Rosengren 1975, Rosengren and Johansson 1976) change valence under pressure their high-pressure hcp phase is more appropriately thought of as belonging to the rare earth sequence, see Fig. 1, whereby it follows that Eu and Yb at very high pressures should exhibit the well-known hcp→Sm-type→dhcp→fcc transitions.

5.4. *The transition metals*

The calculated structural energy differences for the 3d, 4d, and 5d transition metals are shown in Fig. 14 and, as a comparison will show, the predicted crystal structures of all the metals included in this figure, neglecting the three ferromagnetic 3d metals, agree with the experimentally observed crystal structures, Fig. 1, except for the case of Au where the bcc structure is calculated to be marginally more stable than fcc. Hence, it follows that by including complete, i.e. fully hybridized, band structures for each individual metal but still retaining the force theorem one has cured most of the problems connected with the simple canonical picture discussed in Sect. 1.2 and exemplified in Fig. 7. Furthermore, one should note that the correlation between crystal structure and d occupation which the canonical description predicts remains valid also for the complete calculations.

The results in Fig. 14 are very similar to those obtained by Pettifor (1970, 1972, 1977) for the 3d metals and by Williams (quoted by Miedema and Niessen 1983) for the 4d metals. However, in spite of the agreement of the theoretical calculations to within 25% and the correct prediction by the theory of the crystal structures of 27 metals, the calculated structural energy differences are found to be as much as a factor of 3-5 larger than the enthalpy differences obtained from the study of binary phase diagrams (Miedema and Niessen 1983), Fig. 15. At present the cause of this discrepancy is not known. The most likely candidates are either neglect of non-spherical terms in the charge density or a genuine failure of the local density approximation. The force theorem itself does not seem to be the cause of the discrepancy since Williams as quoted by Miedema and Niessen (1983) obtains results similar to ours by subtraction of total energy calculations. Finally, the »experimental« results de-

Fig. 14. Structural energy differences for the 3d, 4d, and 5d transition metals calculated at the experimentally observed equilibrium volume and plotted as functions of the d occupation numbers. The calculations included s, p, and d orbitals but not the Madelung correction Eq. (11).

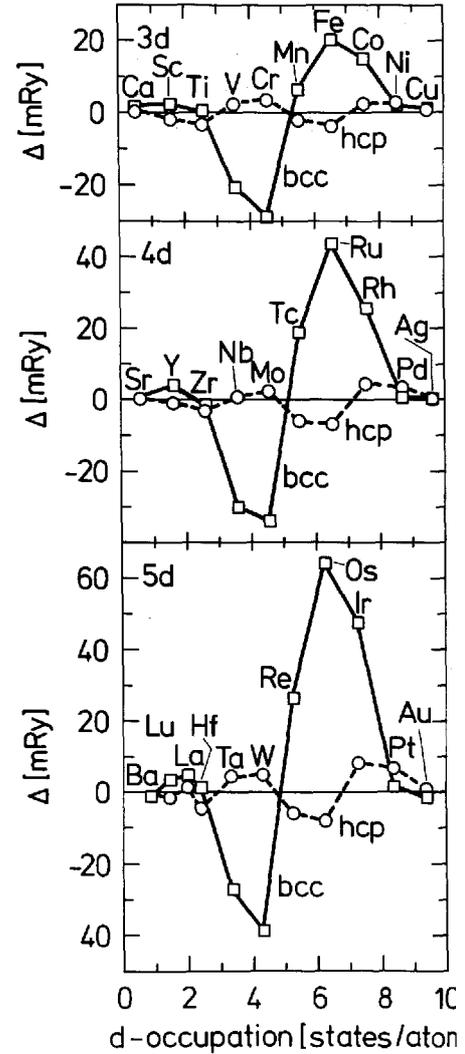
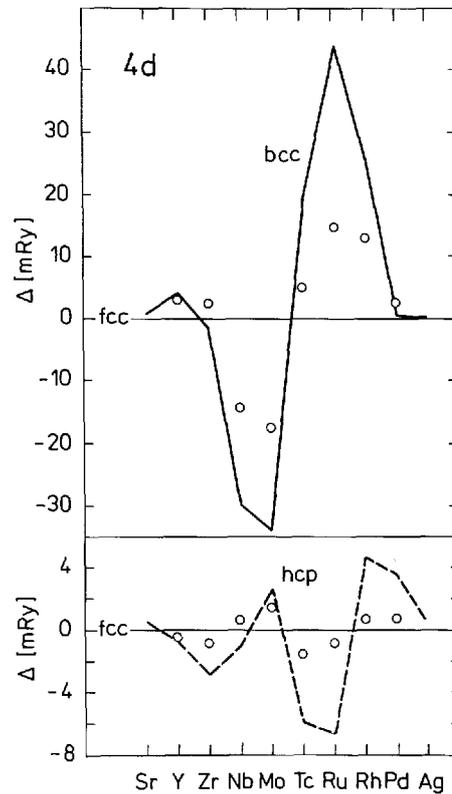


Fig. 15. The calculated bcc-fcc and hcp-fcc structural energy differences (solid and broken lines) for the 4d metals compared with the enthalpy differences derived from phase diagrams (Miedema and Niessen 1983), open circles.

rived by Miedema and Niessen (1983) are certainly model dependent and may therefore have large error bars.

5.5. The lanthanide metals

The calculated structural-energy differences for the two lanthanide metals La and Lu which bracket the lanthanide series are shown in Fig. 16. To compare directly with the canonical results, Fig. 6, the energy differences have been brought onto the canonical scale and plotted as functions of the calculated d occupation number. The results in Fig. 16 are qualitatively similar to the canonical results but the energy differences are generally smaller by approximately a factor of 2, judged by, for instance, the minimum in the Sm-type curve, than their canonical counterparts. Furthermore, the lanthanide sequence has been shifted to lower d occupation numbers whereby the problems connected with the canonical description in the d occupation range above 1.6 have been removed. Hence, Ti, Zr, and Hf are no longer part of the lanthanide sequence and are instead correctly predicted to form in the hcp structure, Fig. 14.

In an account of the cohesive properties of the lanthanides Skriver (1983) found that the d occupation numbers calculated at the experimentally observed equilibrium volume decreased approximately linearly

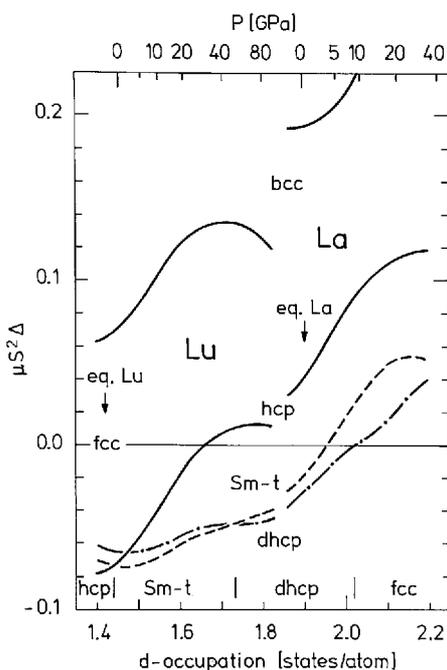
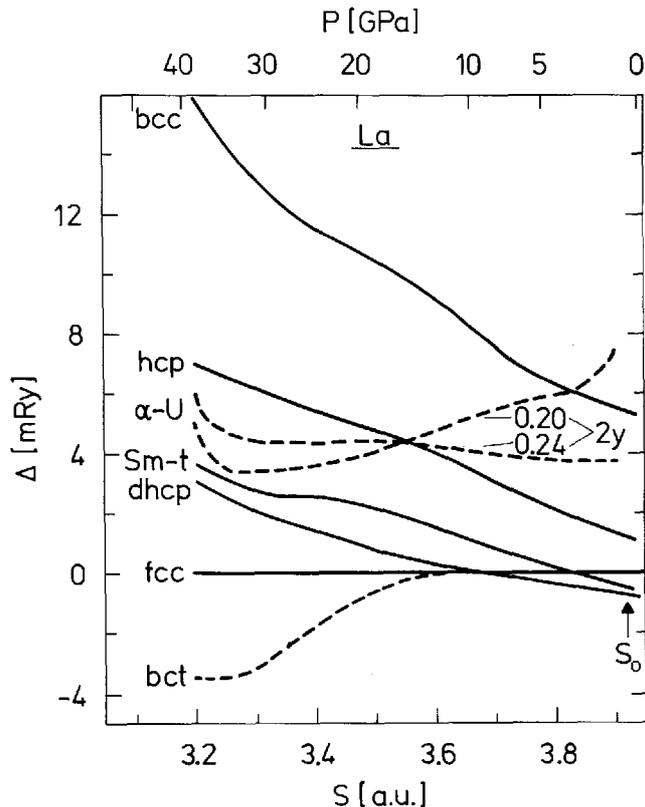


Fig. 16. Structural energy differences for La and Lu calculated as functions of pressure P and plotted versus d occupation number n_d . The calculations included s , p , d , and f orbitals, $4f$ for La and $5f$ for Lu, but not the Madelung correction Eq. (11).

Fig. 17. Structural energy differences for La calculated as functions of pressure P and plotted versus atomic radius. The equilibrium radius $S_0 = 3.92$ [a.u.].



with atomic number between La and Lu. Hence, Fig. 16 may be used to estimate the equilibrium crystal structures of the lanthanide metals, excluding Ce because of its γ - α transition, and the two divalent metals Eu and Yb. In agreement with the generalized phase diagram (Johansson and Rosengren 1975) we find that La, Pr, Nd, and Pm should form in the dhcp structure while Sm should be Sm-type. However, the heavy lanthanides are incorrectly estimated to form in the Sm-type structure. The immediate reason for this failure seems to be that the stability of the hcp structure at a given d occupation is calculated to be too low compared with dhcp and Sm-type but the deeper cause is not known at present. As a result, the Sm-type structure extends over too wide a d occupation range.

Fig. 16 may also be used to predict the behaviour of La and Lu under pressure. We find that Lu should transform from hcp to the Sm-type structure at -2 [GPa] and into the dhcp structure at 35 [GPa]. Because of a 2% error in the calculated equilibrium radius and because of the failure

mentioned above, the first estimate is in error by 25 [GPa], the experimental critical pressure being 23 [GPa] (Liu 1975). The second transition has not yet been observed.

Under pressure La is predicted to transform from dhcp to the fcc structure at 8 [GPa], Fig. 17, which compares favourably with the experimental room-temperature transition pressure of 2.5 [GPa] (Piermarini 1964). The distorted fcc phase discovered by Grosshans et al. (1982) has not been considered, but we shall return to the high-pressure properties of La in the following section.

5.6. Cerium metal under pressure

The behaviour of Ce under pressure has been a subject of long-standing and some controversy, primarily because of the unusual isostructural

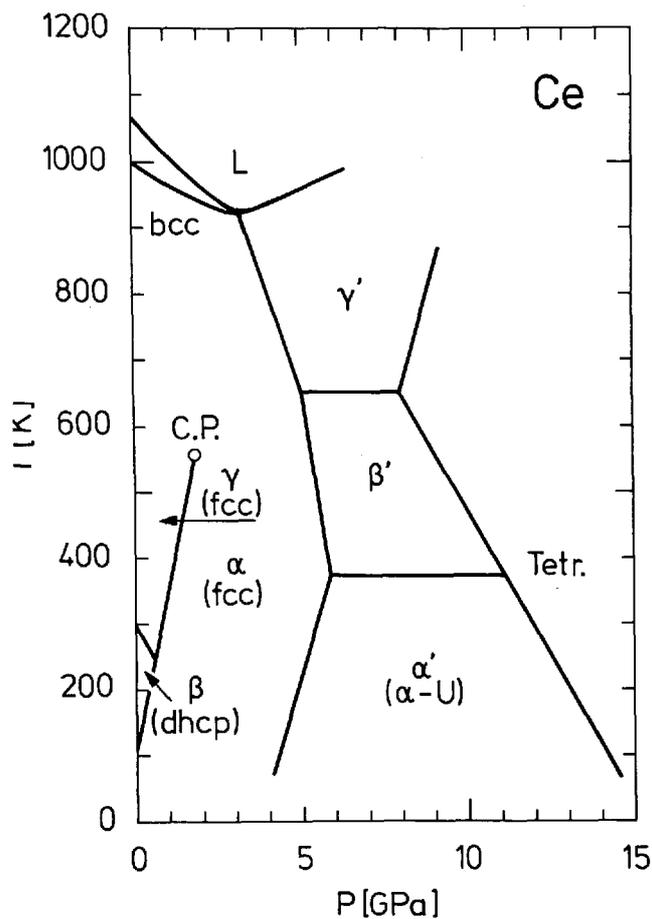


Fig. 18. Phase diagram for Ce compiled from Khvos-tantsev et al. (1983).

$\gamma \rightarrow \alpha$ transition. Here we shall be concerned with the fcc \rightarrow α -U \rightarrow tetragonal crystal-structure sequence exhibited by metallic Ce at low temperature in the pressure range up to 20 [GPa] (see Fig. 18). In the calculations we shall treat the s, p, d, and the 4f electrons on the same footing, i.e. as band electrons. Hence, we favour the picture of the $\gamma \rightarrow \alpha$ transition suggested by Gustafson et al. (1969) and elaborated by Johansson (1974) according to which pressure induces a Mott transition within the 4f shell such that the 4f electron goes from a localized state in γ -Ce to a delocalized, i.e. band state, in α -Ce.

According to the Mott-transition picture Ce metal at pressures above the $\gamma \rightarrow \alpha$ transition is different from the other lanthanides (and indeed from all the other metals we have considered so far) in that it has a fourth conduction electron residing in the 4f band. It is this occupation of the 4f band which is expected to be responsible for the stability of the α -U structure found experimentally above 5.6 [GPa] (Ellinger and Zachariasen 1974) and perhaps for the tetragonal phase found above 12.1 [GPa] (Endo et al. 1977). To shed light on this question we shall now present a series of calculations of structural stabilities for Ce under pressure, and compare the results with those obtained for La where the 4f band is essentially empty.

The orthorhombic α -U structure may be viewed as distorted fcc, where some of the face-centered atoms have been moved away from their positions as described by the parameter $2y$, see Fig. 19. If $2y = 0.5a$ and $a = b = c$ one has the usual fcc unit cell. In the case of Ce the Madelung contribution to the structural energy favours a $2y$ of approximately 0.3 (see top panel of Fig. 19) but the one-electron contribution moves the minimum in the energy difference to $2y = 0.21$ which is the $2y$ value found experimentally in U (Donohue 1975). Under pressure the minimum is seen to move to slightly lower $2y$ values and eventually the α -U structure becomes more stable than the fcc.

From fig. 19 it is expected that Ce will exhibit an fcc \rightarrow α -U phase transition at a pressure which is calculated to be 11.7 [GPa]. The experimental transition pressure is 5.6 [GPa] (Ellinger and Zachariasen 1974), and the discrepancy may be attributed to the fact that the atomic sphere approximation is less suited for open crystal structures such as the α -U structure. As may be seen in Fig. 19 the Madelung correction, which we could neglect for the close-packed crystal structures of the alkaline earth and transition metals is now of the same order of magnitude as the one-electron contribution. Hence, inadequacies in the Madelung approximation of the electrostatic contribution to the structural energy are magni-

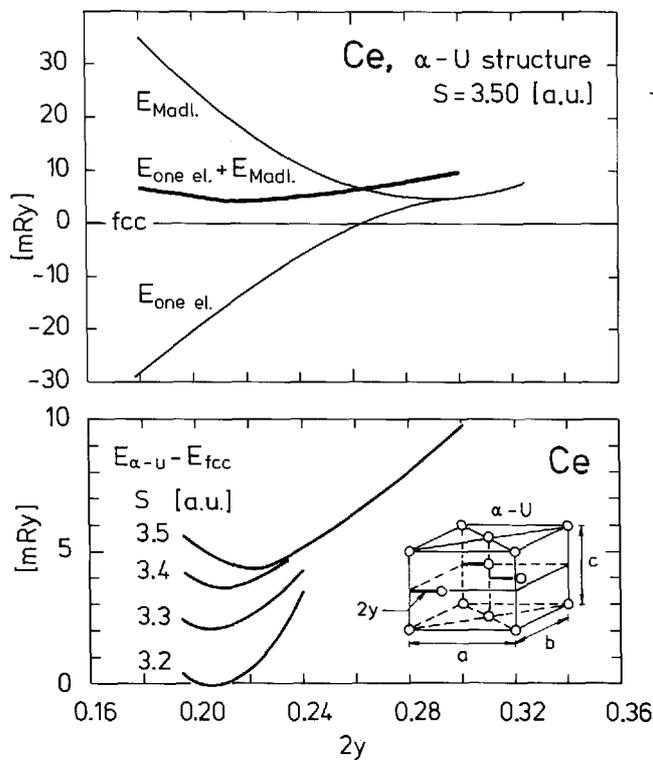


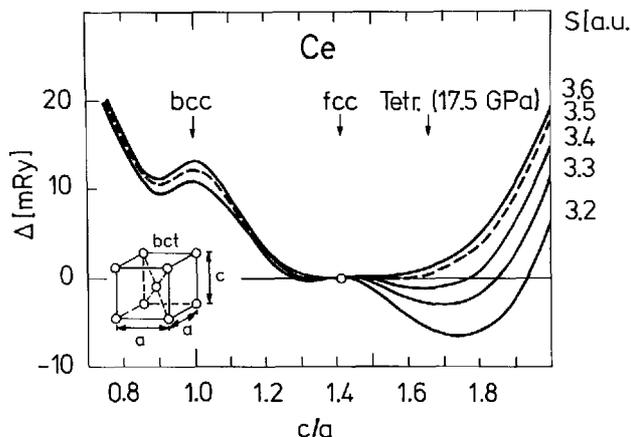
Fig. 19. Energy of Ce in the α -U structure relative to the fcc phase calculated as a function of the positional parameter $2y$ (see insert) and atomic radius S . The individual Madelung and one-electron contributions for one particular radius are shown in the upper panel.

fied and lead to errors in the estimate of the stability of the α -U structure. A similar problem was recently encountered in the case of the open Cs IV structure in Cs metal (McMahan 1984).

If we compare the structural energy-differences for Ce and La (Figs. 17, 19) under pressure we find that while the α -U structure eventually becomes more stable than fcc in Ce it does not do so in La. Since the 4f band is essentially unoccupied in La, whereas Ce has approximately one 4f band electron, the notion that f-band states are responsible for the stability of distorted crystal structures such as the α -U structure is strongly supported by the present calculations. It follows that the α -U structure would not become stable in Ce under pressure unless the 4f electrons were delocalized, i.e. band like, and therefore any adequate description of the α and α' phases in Ce must treat the 4f states on the same footing as the s, p, and d states. In short, Ce is a 4f band metal.

The high-pressure tetragonal structure (Endo et al. 1977) of Ce may be regarded as a distorted fcc structure in which the unit cell has been elongated along the c axis such that the c/a ratio in a body-centred

Fig. 20. Energy of Ce in the body-centred tetragonal (bct) structure relative to the fcc phase calculated as a function of the c/a ratio and atomic radius (the pressures can be inferred from Fig. 21). The insert shows the bct structure.



tetragonal (bct) description is approximately 1.7, see Fig. 20. In the same description bcc and fcc correspond to c/a equal to 1 and $\sqrt{2}$, respectively. According to the structural energy differences in Fig. 20 Ce should as a function of pressure start out in the fcc structure and then transform into a bct structure with a c/a ratio which increases with pressure. In this case the 4f states do not seem to be responsible for the pressure-induced transition, since the same bct structure is also calculated to be the stable high-pressure phase of La, Fig. 17.

In Fig. 21 we have collected the calculated structural energy differences for Ce under pressure. Owing to the less accurate description of open structures discussed above, the α -U structure is seen not to be the stable phase in the pressure range considered, and instead Ce would be expected to go directly from the fcc into the bct phase. However, if we move the α -U curve down by 4.5 [mRy] which is 20% of the Madelung correction (see Fig. 19) we obtain agreement with experiment (Ellinger and Zachariassen 1974, Endo et al. 1977) in the sense that Ce is now expected to exhibit the crystal structure sequence $\text{fcc} \rightarrow \alpha\text{-U} \rightarrow \text{tetragonal}$.

5.7. The light actinides

The calculated structural energy differences for the light actinides Th-Pu are shown in Fig. 22, from which we deduce the most stable close-packed structure to be fcc in Th and Pa and bcc in U, Np, and Pu. This indicates that although these structures are not the stable low-temperature structures in Pa-Pu, they are at least close in energy to the distorted structures observed experimentally and may therefore be realized at elevated temperatures. Experimentally one finds the fcc structure to be

stable in Th up to 1670 [K] (Donohue 1975, Young 1975), and there are indications that Pa has a high temperature fcc phase (Donohue 1975). Furthermore, neither U nor Np has a high temperature fcc phase but instead they become bcc before melting. Pu has a high temperature fcc (δ) phase but since this phase becomes unstable at a pressure of only 0.1 [GPa] it is most probably associated with a localization of the 5f electrons, and the relevant high temperature phase in the present context is then the bcc (ϵ) phase. Thus, experimentally the most stable close-packed structure appears to be fcc in Th and Pa, and bcc in U, Np, and Pu, in agreement with the findings in Fig. 22.

The low-temperature tetragonal structure (α) in Pa may be viewed (Zachariassen 1952) as a distorted bcc structure in which the unit cell has been compressed along the c axis such that the c/a ratio is approximately 0.82, see Fig. 23. According to Fig. 23 the Madelung contribution

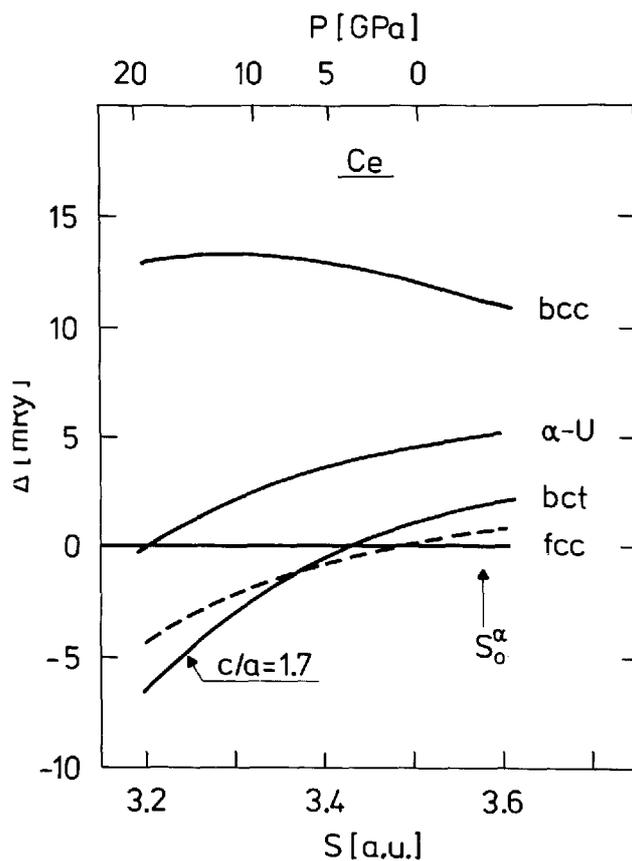


Fig. 21. Structural energy differences for Ce calculated as a function of pressure P and plotted versus atomic radius. S_0^α indicates the experimentally observed equilibrium radius of Ce in the α phase. The calculations included s , p , d , and f orbitals and the Madelung correction Eq. (11).

favours bct structures with c/a in the range from 0.95 to 1.50, whereas structures with c/a outside this range rapidly become extremely unstable. In contrast, the one-electron contribution tends to favour c/a outside the central range, and as a result the energy difference curve for Th has one minimum at $c/a = \sqrt{2}$, corresponding to fcc, in agreement with experiment, while that of Pa exhibits three minima, one of which is close to the c/a observed experimentally in the α phase.

As in the case of the α -U structure in Ce, we are again experiencing problems stemming from the atomic sphere approximation and in particular the Madelung correction, which leads to slightly incorrect estimates of the structural energy differences for open crystal structures. Thus, in the case of Pa the most stable structure is calculated to be bct with $c/a = 1.6$, which incidentally is the high-pressure phase of Ce, whereas the minimum which corresponds to the experimental α structure lies 1.3 mRy above the absolute minimum and is shifted to a c/a of 0.92. However, in view of the rapidly changing Madelung correction in the range below $c/a = 0.95$, it is not unlikely that a better calculation of the electrostatic contribution to the structural energy differences may correct both errors.

Since the 5f band is unoccupied in Th while Pa has approximately one 5f electron it follows from Fig. 23 that the 5f states are responsible for the stability of the tetragonal α phase in Pa. Thus, the situation here is very

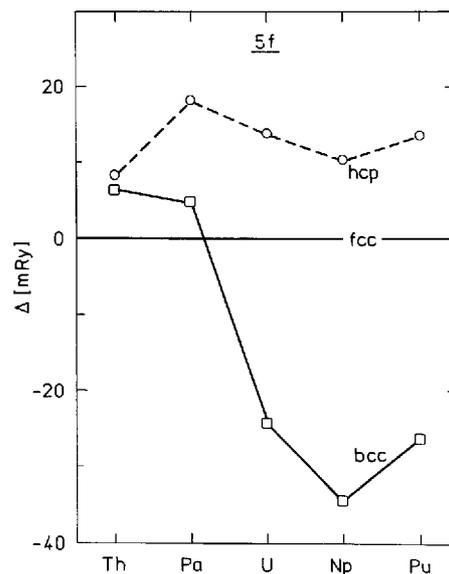


Fig. 22. Calculated structural energy differences for the light actinides plotted versus atomic number. The calculations included s , p , d , and f orbitals but not the Madelung correction Eq. (11).

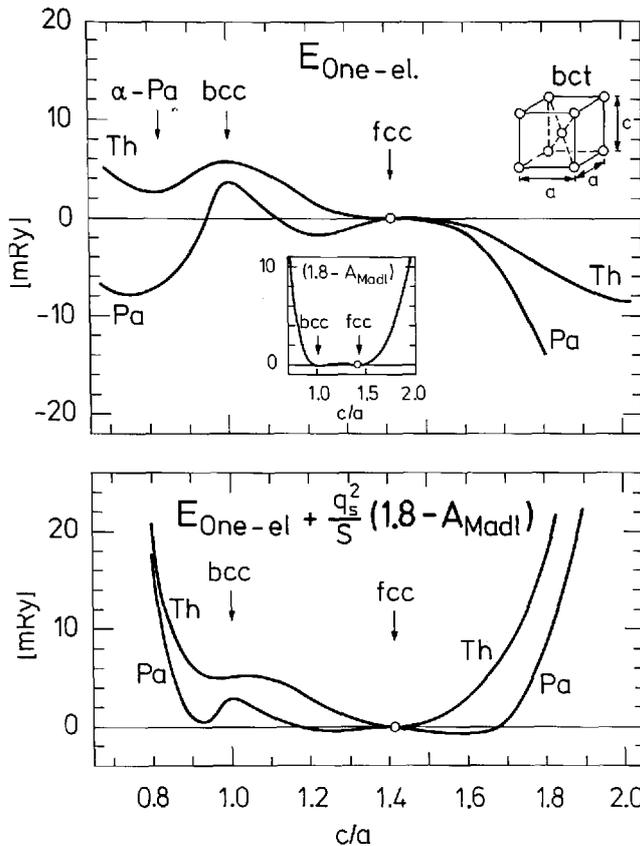


Fig. 23. Energy of Th and Pa in the bct structure relative to the fcc phase calculated as a function of the c/a ratio. The upper panel shows the one-electron contributions, the insert shows the shape of the Madelung correction, and the lower panel shows the total energy differences.

similar to that found earlier in Ce where the presence of one 4f electron stabilized the high pressure α -U structure, and again we take this to mean that the 5f states in the light actinides are itinerant, i.e. band-like, and give rise to distorted crystal structures.

6. Conclusion

We have studied the stability of the crystal structures of some 40 elemental metals within a one-electron approach. The effective one-electron equations have been solved self-consistently by means of the LMTO method and the structural energy differences calculated by means of Andersen's force theorem. This approach has the advantage of treating s, p, d, and f states on the same footing, thus leading to a conceptually

consistent description of trends throughout the periodic table. However, the present implementation of the method is only accurate for close-packed crystal structures, and for that reason we exclude in our study open structures such as CsIV and the more exotic structures found in the actinide series. On the other hand, this shortcoming is not fundamental and will undoubtedly be remedied in the near future.

We find that the theory correctly predicts the crystal structures observed experimentally at low temperature and atmospheric pressure in 35 out of the 42 cases studied. In those few instances where the theory fails we find that the correct crystal structure is only marginally less stable than the calculated structure – this is the case for Na, Au, Yb, and Pa – or the metal is magnetic at low temperature, as in Mn, Fe, and Co. For the light actinides U, Np, and Pu we have not considered the experimentally most stable crystal structures but only the most stable close-packed structures and find the predictions of the theory to be in qualitative agreement with the known phase diagrams.

In a comparison between the calculated structural energy differences for the 4d transition metals and the enthalpy differences derived from studies of phase diagrams we find that, although the crystal structures are correctly predicted by the theory, the theoretical energy differences are up to a factor of 5 larger than their »experimental« counterparts. The reasons for this discrepancy may lie in the local-density approximation or in the neglect of the non-spherical part of the charge distribution. Furthermore, the derived enthalpy differences are certainly model dependent and may change as the model is improved.

In addition to the equilibrium properties we have studied the crystal structures of the alkali, the alkaline earth and some rare earth metals under pressure. We find that the heavy alkalis K, Rb, and Cs should be part of the crystal structure sequence $bcc \rightarrow hcp \rightarrow fcc$ where the intermediate hcp phase may be suppressed at room temperature, and explain the experimentally observed $bcc \rightarrow fcc$ transition in terms of the pressure-induced descent of a zone-boundary energy gap which exists in the fcc band structure but has no counterpart in the bcc case. For the alkaline earth and rare earth metals we find crystal structure sequences which correlate with the calculated d-occupation numbers and which are in agreement with experimental high-pressure observations if we neglect some complex structures found in Ca and Sr.

Finally, we have studied the high-pressure crystal structure sequence $fcc \rightarrow \alpha\text{-U} \rightarrow tet$ for La and Ce and find that under compression the $\alpha\text{-U}$ structure becomes more stable than fcc in Ce, but not in La. This indi-

cates that the presence of itinerant 4f states is responsible for the fcc $\rightarrow\alpha$ -U transition observed experimentally in Ce. In both La and Ce the calculations predict a tetragonal high-pressure phase. This phase is seen experimentally in Ce but not in La where one instead observes a distorted fcc structure not considered in the present work.

In conclusion, we have studied the stability of crystal structures of metals both at equilibrium and at high pressures by a one-electron approach. We find that we can account for the occurrence of most of the close-packed structures observed experimentally. In the few cases where the theory is in disagreement with experiment we find that the correct crystal structure is only marginally less stable than the predicted structure. In order to describe open structures, such as α -U or CsIV, with the same accuracy as the close-packed structures one needs a more accurate approximation for the electrostatic contribution to the total energy.

ACKNOWLEDGMENTS. The present series of calculations grew out of conversations with several people. It is thus a great pleasure to thank K. Syassen and K. Takemura for making me interested in the alkali metal problem, and B. Johansson for suggesting the Ce problem. B. Johansson and A. K. McMahan have furthermore helped clarify calculational as well as experimental problems. Part of this work was started while visiting Los Alamos Scientific Laboratory, and I wish to thank the group at the Materials Science Center for its kind hospitality. Finally, I want to express my gratitude to Knud Højgaard's Fond for granting my Niels Bohr Fellowship.

References

- J. Akella, Q. Johnson, W. Thayer, and R. N. Schock, *J. Less-Comm. Met.* **68**, 95 (1979).
J. Akella, Q. Johnson, and R. N. Schock, *Geophysical Res.* **85**, 7056 (1980).
I. V. Alexandrov, C. V. Nesper, V. N. Katchinsky and J. Marenko, paper at the 20th meeting of the »European High pressure Research Group«, Stuttgart, 1982 (unpublished).
O. K. Andersen, *Phys. Rev.* **B12**, 3060 (1975).
O. K. Andersen and O. Jepsen, *Physica* **91B**, 317 (1977).
O. K. Andersen, J. Madsen, U. K. Poulsen, O. Jepsen, and J. Kollar, *Physica* **86-88B**, 249 (1977).
O. K. Andersen, H. L. Skriver, H. Nohl, and B. Johansson, *Pure Appl. Chem.* **52**, 93 (1979).
A. O. E. Animalu, *Phys. Rev.* **161**, 445 (1967).
J. Bardeen, *J. Chem. Phys.* **6**, 372 (1938).
U. von Barth and L. Hedin, *J. Phys.* **C5**, 1629 (1972).
B. J. Beaudry and K. A. Gschneidner, Jr., »Preparation and Basic Properties of the Rare Earth Metals« in »*Handbook on the Physics and Chemistry of Rare Earths*« ed. by K. A. Gschneidner, Jr. and L. R. Eyring (North Holland Publishing Company, Amsterdam 1978).

- U. Benedict, J. R. Peterson, R. G. Haire, and C. Dufour, *J. Phys.* *F14*, L43 (1984).
- M. Born and J. E. Mayer, *Zeits. f. Physik* *75*, 1 (1932).
- L. Brewer: »Phase Stability in Metals and Alloys« in »*Batelle Institute Materials Science Colloquia*«, ed. by P. S. Rudman, J. Stringer, and R. I. Jaffee (McGraw-Hill, New York 1967), pp. 39-62.
- E. Bucher, P. H. Schmidt, A. Jayaraman, K. Andres, J. P. Maita, K. Nassau, and P. D. Dernier, *Phys. Rev.* *B2*, 3911 (1970).
- M. S. S. Brooks and B. Johansson, *J. Phys* *F13*, L197 (1983).
- N. W. Dalton and R. A. Deegan, *J. Phys.* *C2*, 2369 (1969).
- R. A. Deegan, *J. Phys.* *C1*, 763 (1968).
- J. Donohue, »The Structures of the Elements« (John Wiley & Sons, New York 1975).
- F. Ducastelle and F. Cyrot-Lackmann, *J. Phys. Chem. Solids* *32*, 285 (1971).
- K. J. Dunn and F. P. Bundy, *Phys. Rev.* *B24*, 1643 (1981).
- J. C. Duthie and D. G. Pettifor, *Phys. Rev. Lett.* *38*, 564 (1977).
- F. H. Ellinger and W. H. Zachariassen, *Phys. Rev. Lett.* *32*, 773 (1974).
- S. Endo, H. Sasaki, and T. Mitsui, *J. Phys. Soc. Japan* *42*, 882 (1977).
- E. Esposito, A. E. Carlsson, D. D. Ling, H. Ehrenreich, and C. D. Gelatt, Jr., *Phil. Mag.* *A41*, 251 (1980).
- J. Friedel, »Transition Metals. Electronic Structure of the d-band. Its Role in the Crystalline and Magnetic Structures« in »*The Physics of Metals 1. Electrons*« ed. by J. M. Ziman (Cambridge University Press 1969).
- J. Friedel and C. M. Sayers, *J. Physique* *38*, 697 (1977).
- D. Glötzel and A. K. McMahan, *Phys. Rev.* *B20*, 3210 (1979).
- W. A. Grosshans, Y. K. Vohra, and W. B. Holzapfel, *Phys. Rev. Lett.* *49*, 1572 (1982).
- K. A. Gschneidner, »Physical Properties and Interrelations of Metallic and Semimetallic Elements« in »*Solid State Physics*« vol. *16*, ed. by H. Ehrenreich, F. Seitz, and D. Turnbull (Academic Press, New York 1964).
- K. A. Gschneidner and R. M. Valletta, *Acta Met.* *16*, 477 (1968).
- D. R. Gustafson, J. D. McNutt and L. O. Roellig, *Phys. Rev.* *183*, 435 (1969).
- J. Hafner and V. Heine, *J. Phys.* *F13*, 2479 (1983) and references therein.
- H. T. Hall, L. Merrill, and J. D. Barrett, *Science* *146*, 1297 (1964).
- V. Heine and D. Weaire, »Pseudopotential Theory of Cohesion and Structure« in »*Solid State Physics*« vol. *24* ed. by H. Ehrenreich, F. Seitz, and D. Turnbull (Academic press, New York 1970).
- V. Heine, »Electronic Structure from the Point of View of the Local Atomic Environment« in »*Solid State Physics*« vol. *35* ed. by H. Ehrenreich, F. Seitz, and D. Turnbull (Academic Press, New York 1980).
- C. H. Hodges, *Acta Met.* *15*, 1787 (1967).
- W. B. Holzapfel, T. G. Ramesh, and K. Syassen, *J. de Phys. Coll.* *40*, C5-390 (1979).
- J.-P. Jan and H. L. Skriver, *J. Phys.* *F11*, 805 (1981).
- A. Jayaraman, W. Klement, Jr., and G. C. Kennedy, *Phys. Rev. Lett.* *10*, 387 (1963a).
- A. Jayaraman, N. Klement, Jr., and G. C. Kennedy, *Phys. Rev.* *132*, 1620 (1963b).
- A. Jayaraman, *Phys. Rev.* *135*, A1056 (1964).
- A. Jayaraman and R. C. Sherwood, *Phys. Rev.*, *134*, A691 (1964).
- A. Jayaraman, *Phys. Rev.* *139*, A690 (1965).
- B. Johansson, *Phil. Mag.* *30*, 469 (1974).
- B. Johansson and A. Rosengren, *Phys. Rev.* *B11*, 2836 (1975).

- B. Johansson, »Structural and elastic properties of the f elements« in »*Rare Earths and Actinides, 1977*« ed. by W. D. Corner and B. K. Tanner (IOP, Bristol 1978), p. 39 (Inst. Phys. Conf. Ser. No. 37).
- H. Jones, Prog. Phys. Soc. 49, 250 (1937).
- L. Kaufman and H. Bernstein: »Computer Calculations of Phase Diagrams« (Academic Press, New York 1970).
- L. G. Khvostantsev and N. A. Nikolaev, Phys. Stat. Sol. (a) 77, 161 (1983).
- C. C. Koch, J. Less-Comm. Met. 22, 149 (1970).
- W. Kohn and L. J. Sham, Phys. Rev. 140A, 1135 (1965).
- L.-G. Liu, W. A. Bassett, and M. S. Liu, Science 180, 298 (1973).
- L.-G. Liu, J. Phys. Chem. Solids 36, 31 (1975).
- J. W. McCaffrey, J. R. Anderson, and D. A. Papaconstantopoulos, Phys. Rev. B7, 674 (1973).
- A. R. Machintosh and O. K. Andersen, »The electronic structure of transition metals« in »*Electrons at the Fermi surface*« ed. by M. Springford (Cambridge University Press, Cambridge 1980).
- A. K. McMahan and J. A. Moriarty, Phys. Rev. B27, 3235 (1983).
- A. K. McMahan, Phys. Rev. B29, 5982 (1984).
- D. B. McWhan and A. L. Stevens, Phys. Rev. 139, A682 (1965).
- D. B. McWhan and A. L. Stevens, Phys. Rev. 154, 438 (1967).
- D. B. McWhan, T. M. Rice, and P. H. Schmidt, Phys. Rev. 177, 1063 (1969).
- D. J. Michish, A. B. Kunz, and S. T. Pantalides, Phys. Rev. B10, 1369 (1974).
- A. R. Miedema and A. K. Niessen, CALPHAD 7, 27 (1983).
- J. A. Moriarty, Phys. Rev. B8, 1338 (1973).
- J. A. Moriarty, Phys. Rev. B26, 1754 (1982).
- J. A. Moriarty and A. K. McMahan, Phys. Rev. Lett. 48, 809 (1982).
- N. F. Mott and H. Jones, »The Theory of the Properties of Metals and Alloys« (Oxford Univ. Press, London 1936).
- A. Nakaue, J. Less-Comm. Met. 60, 47 (1978).
- R. M. Nieminen and C. H. Hodges, J. Phys. F6, 573 (1976).
- H. Olijnyk and W. B. Holzapfel, Phys. Lett. 99A, 381 (1983).
- H. Olijnyk and W. B. Holzapfel, Phys. Lett. 100A, 191 (1984).
- B. Olinger and J. W. Shaner, Science 219, 1071 (1983).
- D. G. Pettifor, J. Phys. C3, 367 (1970).
- D. G. Pettifor: »Theory of the Crystal Structures of Transition Metals at Absolute Zero« in »*Metallurgical Chemistry*« ed. by O. Kubashewski (HMSO, London 1972).
- D. G. Pettifor, Commun. Phys. 1, 141 (1976).
- D. G. Pettifor, CALPHAD 1, 305 (1977).
- G. J. Piermarini and C. E. Weir, Science 144, 69 (1964).
- R. B. Roof, R. G. Haire, D. Schiferl, L. A. Schwalbe, E. A. Kmetko, and J. L. Smith, Science 207, 1353 (1980).
- R. B. Roof, Z. für Krist. 158, 307 (1982).
- A. Rosengren and B. Johansson, Phys. Rev. B13, 1468 (1976).
- H. L. Skriver, Phys. Rev. Lett. 49, 1768 (1982).
- H. L. Skriver, »Electronic Structure and Cohesion in The Rare Earth Metals« in »*Systematics and the Properties of the Lanthanides*« ed. S. P. Sinha (D. Reidel Publishing Company, Dordrecht 1983).

- H. L. Skriver, »The LMTO Method« (Springer-Verlag, Berlin 1984).
- P. C. Souers and G. Jura, *Science* *140*, 481 (1963).
- R. A. Stager and H. G. Drickamer, *Phys. Rev.* *131*, 2524 (1963).
- R. A. Stager and H. G. Drickamer, *Science* *139*, 1284 (1963).
- D. R. Stephens, H. D. Stromberg, and E. M. Lilley, *J. Phys. Chem. Solids* *29*, 815 (1968).
- K. Syassen, G. Wortmann, J. Feldhaus, K. H. Frank, and G. Kaindl, *Phys. Rev.* *B26*, 4745 (1982).
- K. Takemura, S. Minomura, and O. Shimomura, »Structure of Cesium and Iodine under Pressure« in »*Physics of Solids under High Pressure*« ed. by J. S. Schilling and R. N. Shelton (North-Holland, Amsterdam 1981).
- K. Takemura and K. Syassen, *Solid State Commun.* *44*, 1161 (1982).
- K. Takemura, S. Minomura, and O. Shimomura, *phys. Rev. Lett.* *49*, 1772 (1982).
- K. Takemura and K. Syassen, *Phys. Rev.* *B28*, 1193 (1983).
- K. Takemura and K. Syassen, *J. Phys. F* (1985, in print).
- A. R. Williams, unpublished, and quoted by Miedema and Niessen (1983).
- Y. K. Vohra, H. Olijnyk, W. A. Grosshans, and W. B. Holzapfel, *Phys. Rev. Lett.* *47*, 1065 (1981).
- D. A. Young, »Phase Diagrams of the Elements«, Lawrence Livermore Laboratory Report UCRL-51902 (1975, unpublished).
- D. A. Young and M. Ross, *Phys. Rev.* *B29*, 682 (1984).
- W. H. Zachariasen, *Acta Cryst.* *5*, 19 (1952).
- A. Zunger and M. L. Cohen, *Phys. Rev.* *B18*, 5449 (1978); *B20*, 4082 (1979).

CLAUS S. JACOBSEN

Infrared Studies on the Electronic Structure of Organic Conductors

ABSTRACT. *It is experimentally shown that a simple Drude analysis of the reflectance edge in organic conductors yields reliable values for electronic transfer integrals. The integrated infrared oscillator strength is found to be smaller than expected from band theory. The reduction is interpreted as being an effect of the short range electron-electron interaction and is used to compare the strength of this interaction among different materials. It is concluded that highly correlated systems may be semiconductors due to electron-phonon driven instabilities, or, if the interstack coupling is sufficient, may remain good metals to low temperatures. The role of the electron-molecular vibration coupling is stressed, both as contributing to instabilities, and as a microscopic probe.*

Physics Laboratory 3, Technical University of Denmark,
DK-2800 Lyngby, Denmark.

Introduction

An ordinary metal, like copper, is characterized by two basic features: (1) A high concentration of valence electrons, which are in principle free to move through the metal, and (2) a strong overlap of the valence electron orbitals (wavefunctions) on neighbour atoms, which effectively makes the electrons delocalize in the metal. The only force capable of inhibiting the free motion of the electrons is the direct Coulomb repulsion between the negatively charged particles. If strong enough, such a repulsion would localize the electrons, one on each atom (Mott, 1949). Although the arguments are quite complicated, it is now well understood, how the electronic system itself almost completely screens out the Coulomb repulsion. The effectiveness of the screening is due to the same two features listed above: A high carrier density and a good neighbour contact. As a consequence ordinary metals have high electrical conductivities (e.g. copper at room temperature with $\sigma = 6 \times 10^5 \text{ Scm}^{-1}$).

Furthermore, the conductivity increases when the metal is cooled, since the scattering rate for the electrons decreases as the thermally induced vibrations of the atoms get smaller in amplitude. Some metals

even become superconducting at very low temperatures (a few Kelvin), that is, their electrical resistivity vanishes completely.

Such behaviour is contrasted by semiconductors like silicon, where the charge carriers are thermally excited from a state which is insulating at low temperatures. All electrons take part in the localized crystal binding. Thus typically, the conductivity of semiconductors increases with temperature.

In a recently very active, interdisciplinary field of the materials sciences, chemists and physicists have attempted to mimic metallic, electrical properties in organic compounds. The long term perspective is to make it possible employing the great flexibility of organic chemistry to design, synthesize, and manufacture materials with specific, desirable properties.

Presently, there are two approaches to the synthetic metal problem. A considerable amount of effort goes into attempts of doping to high levels polymeric semiconductors like polyacetylene, $(\text{CH})_x$ (for a review see, for example, Baeriswyl et al. (1982)). Although promising for applications, such materials are not truly metallic, since their carrier concentrations are fairly low.

The other approach has led to the large class of materials, normally called organic conductors (see, for example, Jerome and Schulz (1982), and references therein). Here the work has been directly aimed at incorporating the two basic features of a metal into crystals of organic molecules. The high concentration of charge carriers is obtained by bringing together at least two species of molecules, one which is willing to accept an extra, unpaired electron (acceptor molecule), and another which readily gives up an electron (donor molecule). Thus in the crystal, charge is transferred from donors to acceptors. The second feature, good contact between neighbour building blocks, i.e. the molecules, is achieved by using near planar, π -bound organic molecules, which tend to form stacks with a fair inside overlap. It follows that the contact between stacks must be rather weak, hence the materials are electrically highly anisotropic. In some contexts, they may even be designated one-dimensional (1D).

Thus to shortly characterize organic conductors, they are synthetic materials, which by elementary solid state physics are expected to be metals in electrical sense. However, they are highly anisotropic, have carrier densities one or two orders of magnitude below those of ordinary metals, and even along the molecular stacks, the neighbour contact is at least ten times weaker than in, for example, copper. The latter features

arise from the use of rather big building blocks, which do not pack as effectively as compact metal atoms.

Therefore, it is not surprising that organic conductors so far have been found to be inferior to the elements with respect to metallic quality. The first organic conductors were in fact semiconducting. These early studies date back to the 1950s and -60s (see Gutmann and Lyons, 1967). In 1973 the first material with genuine metallic properties was synthesized (Ferraris et al., 1973). TTF-TCNQ has a room temperature conductivity of order 600 Scm^{-1} , and a strongly increasing $\sigma(T)$ down to 60K. However, below 60K a metal-insulator transition destroys the conductivity. In the following years many derivatives of TTF-TCNQ were studied, and in 1980 Jerome et al. reported superconductivity in the material $\text{TMTSF}_2\text{PF}_6$ (below 1K and only under a pressure of ~ 10 kbar). The present state of art involves several ambient pressure superconductors (working up to 2.5K), as well as materials with a wide range of interesting magnetic and non-magnetic ground states. Many of the metal-insulator and magnetic phase transitions may be related to the low-dimensionality of the materials. Thus while applications are still somewhat in the future, the materials constitute extremely interesting model systems for solid state physics.

It is the aim of the present study to use infrared (IR) spectroscopy in characterizing the basic interactions in organic conductors. We shall discuss how the IR properties give direct information on the intra- as well as interstack contacts, on the interplay between electron-electron and electron-vibration interactions, and to some extent on the detailed nature of the phase transitions occurring in these materials. It is a central result of the study, that the short range electron-electron interaction plays a role far more important than in ordinary metals. It indeed appears that the probability for finding two conduction electrons on the same molecule is quite small in several materials. The electrons then more or less behave as spinless fermions in transport and optical properties.

Although essential for a microscopic understanding of the electronic structure, the electron-electron interaction needs not destroy the metallic state. Even the organic superconductors appear to have a sizeable electron-electron Coulomb interaction, but a large interchain overlap reduces its impact on the physical properties.

Materials and methods

All the materials investigated in this study belong to the group of potential metals, which are characterized by partially filled one-electron bands and no strong static disorder. The constituent organic molecules are shown in Fig. 1, and their systematic names are listed in Table I.

We will deal with two groups of compounds. (1) The doublestack conductors, like TTF-TCNQ, have uniform stacks, and donor as well as acceptor chains have unpaired electrons. The crystal structure is sketched in Fig. 2(a). A key parameter for these materials is the degree of charge transfer, q , equal to the average number of carriers per molecule. q is determined by a complicated energy balance (Torrance, 1979), and is best found indirectly by diffuse X-ray scattering, which detects weak superstructures related to the Fermi wavevector. (2) The singlestack conductors, dealt with here, are complex salts, where the stoichiometry is such that there is one closed shell ion for a pair of organic molecules. The counterion may be organic or inorganic, but q for the conducting

Table I. Organic molecules. D = donor. A = acceptor

Abbr.	Type	Systematic name
TTF	D	$\Delta^{2,2'}$ -bi-1,3-dithiolylidene
TMTTF	D	$\Delta^{2,2'}$ -bi-4,5-dimethyl-1,3-dithiolylidene
DBTTF	D	$\Delta^{2,2'}$ -bibenzo-1,3-dithiolylidene
HMTTF	D	$\Delta^{2,2'}$ -bicyclopenteno-1,3-dithiolylidene
TSF	D	$\Delta^{2,2'}$ -bi-1,3-diselenolylidene
TMTSF	D	$\Delta^{2,2'}$ -bi-4,5-dimethyl-1,3-diselenolylidene
DBTSF	D	$\Delta^{2,2'}$ -bibenzo-1,3-diselenolylidene
HMTSF	D	$\Delta^{2,2'}$ -bicyclopenteno-1,3-diselenolylidene
BEDT-TTF (ET)	D	$\Delta^{2,2'}$ -bi-5,6-dihydro-1,4-dithiino-1,3-dithiolylidene
MEM ⁺	D ^a	N-ethyl-N-methyl-morpholinium
TCNQ	A	7,7,8,8-tetracyano-p-quinodimethane
DMTCNQ	A	2,5-dimethyl-7,7,8,8-tetracyano-p-quinodimethane
TCNQCl ₂	A	2,5-dichloro-7,7,8,8-tetracyano-p-quinodimethane
TCNQF ₄	A	2,3,5,6-tetrafluoro-7,7,8,8-tetracyano-p-quinodimethane
TNAP	A	11,11,12,12-tetracyanonaphtho-2,6-quinodimethane

^a closed shell ion

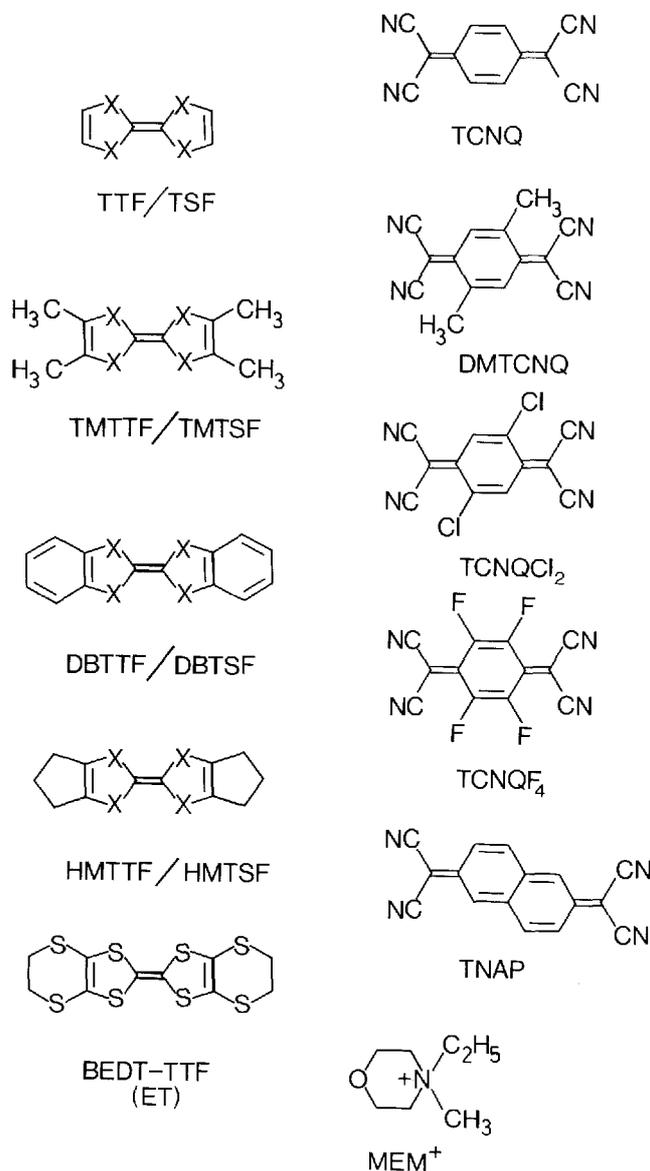


Fig. 1. Donor and acceptor molecules (cfr. Table I).

stack is always 0.5. Due to the stoichiometry alone the stacks show some dimerisation. A typical structure is sketched in Fig. 2(b).

Table II lists the relevant materials with basic physical properties, and gives references to more detailed information.

The organic conductors are usually available as small, 2-4 mm long needles with cross-sectional dimensions of a few tenths of a mm. The

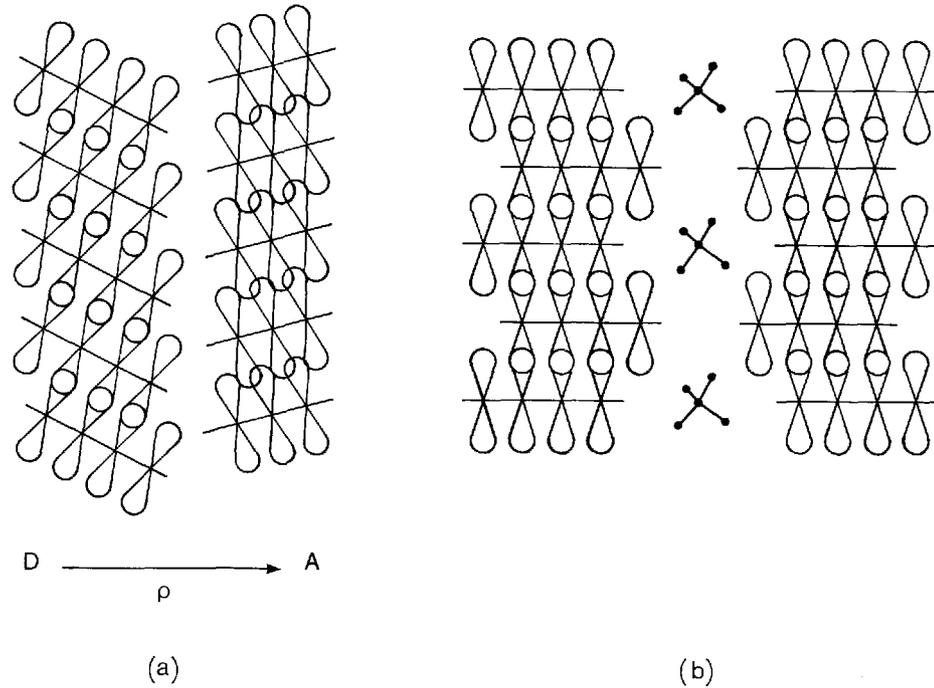


Fig. 2. Schematic crystal structures. (a) Double-stack conductor. ρ is the degree of charge transfer. (b) Single-stack conductor with inactive counterions.

faces tend to be of high optical quality, and since the crystals are opaque in the entire range, studies of the IR properties are best conducted as specular, polarized, near-normal incidence reflectance measurements. Details on experimental equipment and procedures are described by Jacobsen et al. (1983).

A frequency range as broad as possible is covered. Then the Kramers-Kronig relations, valid for linear, causal and local response functions may be employed to extract information on the complex dielectric function, $\tilde{\epsilon}(\omega)$. From the power reflectance, $R(\omega)$, it is possible to calculate the phase shift on reflection, $\theta(\omega)$:

$$\theta(\omega) = \frac{\omega}{\pi} \text{P} \int_0^{\infty} \frac{\ln R(\omega')}{\omega'^2 - \omega^2} d\omega', \quad (1)$$

and then

$$\tilde{\epsilon}(\omega) = \left(\frac{1 + \sqrt{R(\omega)} e^{i\theta(\omega)}}{1 - \sqrt{R(\omega)} e^{i\theta(\omega)}} \right)^2. \quad (2)$$

Table II. Characteristics of some organic conductors. The third column gives high/low temperature character (M = metal, S = semiconductor, SM = semimetal, SC = superconductor, I = insulator).

Material	$\sigma_{dc}(300K)$ S cm ⁻¹	Cha- racter	ρ	Structure
TTF-TCNQ	600 Cohen et al. (1974)	M/S	0.55 Kagoshima et al. (1976)	monoclinic Kistenmacher et al. (1974)
TSF-TCNQ	800 Etemad et al. (1975)	M/S	0.63 Weyl et al. (1976)	monoclinic Etemad et al. (1975)
TMTSF-TCNQ	1000 Jacobsen et al. (1978)	M/S	0.57 Pouget (1981)	triclinic Bechgaard et al. (1977)
TMTTF -DMTCNQ	120 Jacobsen et al. (1978)	M/S	-	triclinic Jacobsen et al. (1978)
TMTSF -DMTCNQ	500 Jacobsen et al. (1978)	M/S	0.50 Pouget (1981)	triclinic Andersen et al. (1978)
HMTTF-TCNQ	400 Greene et al. (1976)	M/S	0.72 Megtert et al. (1978)	orthorhombic Greene et al. (1976)
HMTSF-TCNQ	2000 Bloch et al. (1975)	M/SM	0.74 Weyl et al. (1976)	orthorhombic Phillips et al. (1976)
DBTTF -TCNQCl ₂	40 Jacobsen et al. (1980)	S/I	0.56 Mortensen et al. (1983)	triclinic Soling et al. (1981a)
HMTSF-TNAP	2400 Bechgaard et al. (1978)	M/SM	0.58 Pouget (1984)	triclinic Kistenmacher (1978)
DBTSF -TCNQF ₄	0.0001 Bryden et al. (1984)	S/I	1.0 Bryden et al. (1984)	monoclinic Bryden et al. (1984)
MEM-TCNQ ₂	0.001 Huizinga et al. (1979)	S/I	0.5	triclinic Bosch and v.Bodegom (1977)
TMTSF ₂ PF ₆ , -AsF ₆ , -SbF ₆	500 Bechgaard et al. (1980)	M/S	0.5	triclinic Thorup et al. (1981) Soling et al. (1981b)
TMTSF ₂ ClO ₄	700 Bechgaard et al. (1981a)	M/SC	0.5	triclinic Bechgaard et al. (1981b)
TMTTF ₂ Br	260 Delhaes et al. (1979)	M/I	0.5	triclinic Liautard et al. (1982)
TMTTF ₂ PF ₆	20 Delhaes et al. (1979)	S/I	0.5	triclinic Liautard et al. (1982)
(BEDT-TTF) ₂ I ₃	30 Yagubskii et al. (1984)	M/SC	0.5	triclinic Kaminskii et al. (1984)

P denotes principal value. Clearly in Eq. (1) suitable extrapolations for the ranges not covered must be adopted (see, for example, Wooten (1972)). However, if the measured range is sufficiently wide, $\bar{\epsilon}(\omega)$ is not very dependent on the particular choice of extrapolations.

For conducting substances, the imaginary part of $\bar{\epsilon}(\omega)$ diverges as $\omega \rightarrow 0+$. Hence it is more convenient to introduce the real dielectric function, $\epsilon(\omega)$, and the frequency dependent conductivity, $\sigma(\omega)$, related to $\bar{\epsilon}(\omega)$ by

$$\bar{\epsilon}(\omega) = \epsilon(\omega) + i\sigma(\omega)/\epsilon_0\omega. \quad (3)$$

It is noted that the area under the $\sigma(\omega)$ curve in a natural way corresponds to the optical oscillator strength. For example, integrating over all frequencies yields the sum rule (Wooten, 1972),

$$\int_0^\infty \sigma(\omega)d\omega = (\pi/2)Ne^2/m, \quad (4)$$

where N is the total electron density and m is the electron mass. As demonstrated later, similar sum rules may with some care be applied to limited frequency ranges and particular groups of electrons.

Basic interactions and instabilities

This section introduces important physical parameters and models, which will be referred to later. Theoretical expectations to the IR properties associated with the models are discussed. Finally, we review shortly the different types of low-dimensional instabilities.

One-electron model

The simplest model for the molecular chain compounds assumes non-interacting electrons, which can move only in the chain direction:

$$H = \sum_{i,\sigma} E_0 n_{i,\sigma} - \sum_{i,\sigma} t(c_{i,\sigma}^+ c_{i+1,\sigma} + c_{i+1,\sigma}^+ c_{i,\sigma}). \quad (5)$$

$c_{i,\sigma}^+$ creates an electron of spin projection σ on site i , and $n_{i,\sigma} = c_{i,\sigma}^+ c_{i,\sigma}$ is the occupation number. E_0 is the solid state ionization potential, and t is the transfer integral associated with the finite overlap between near neighbour orbitals. More distant overlaps may be safely ignored in molecular crystals. This tight-binding approximation leads to a cosine energy band:

$$\epsilon_k = E_0 - 2t\cos kd, \quad (6)$$

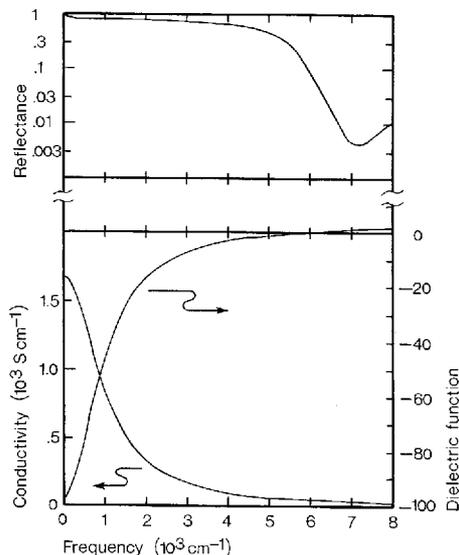


Fig. 3. Optical properties of the Drude model. Notice the logarithmic reflectance scale. Parameters: $\epsilon_c = 3.0$, $\omega_p = 10000 \text{ cm}^{-1}$, $\gamma_{opt} = 1000 \text{ cm}^{-1}$.

where d is the molecular repeat distance. With q electrons per site, the band is filled to $\pm k_F = \pm \pi q / 2d$ (assuming $t > 0$). Thus the Fermi surface consists of the two parallel planes $k = \pm k_F$. The density of states at the Fermi level is $N(\epsilon_F) = (\pi t \sin(\pi q / 2))^{-1}$.

The IR properties associated with Eq. (5) may be derived in the self-consistent field approximation (see, for example, Wooten (1972)). Assuming a frequency independent relaxation rate, γ_{opt} , and a background dielectric constant arising from virtual high frequency transitions, ϵ_c , the result is of the Drude form,

$$\tilde{\epsilon}(\omega) = \epsilon_c - \omega_p^2 / \omega(\omega + i\gamma_{opt}), \quad (7)$$

where the plasma frequency, ω_p , may be calculated from

$$\omega_p^2 = (e^2 / \epsilon_0 \hbar^2) \sum_{BZ} f(\epsilon_k) \partial^2 \epsilon_k / \partial k^2. \quad (8)$$

$f(\epsilon_k)$ is the Fermi-Dirac occupation number, and the derivative $\partial^2 \epsilon_k / \partial k^2$ is to be taken along the direction of the electric field.

In the present quasi-1D model, Eq. (8) yields

$$\omega_p^2 = 4td^2 e^2 \sin(\pi q / 2) / \pi \epsilon_0 \hbar^2 V_m, \quad (9)$$

where V_m is the crystal volume per molecule. This expression is derived for $T = 0$. The explicit temperature dependence may be ignored when $k_B T \ll \epsilon_F$, a condition which is normally fulfilled. In Fig. 3 the IR

properties of the Drude model are shown for typical values of the parameters.

Especially noteworthy is the characteristic drop in reflectance in the near IR, which arises from the zero-crossing of $\epsilon(\omega)$. This is also the frequency of the plasmons, i.e. long wavelength oscillations in the conduction electron charge density. Since the plasmons are longitudinal excitations, they are not optically excitable in a normal incidence experiment, but the sharp drop in reflectance, the plasma edge, is a signature of their existence.

Another point is that the area below the $\sigma(\omega)$ curve is related to ω_p by a partial sum rule

$$\int_{\text{intra band}} \sigma(\omega) d\omega = (\pi/2)\epsilon_0\omega_p^2. \quad (10)$$

Interchain hopping may easily be incorporated into the model. If, for simplicity, we consider an orthorhombic, 2D model with transfer integrals t_{\parallel} and t_{\perp} , the band dispersion is given by:

$$\epsilon_{\vec{k}} = E_0 - 2t_{\parallel}\cos(k_{\parallel}d_{\parallel}) - 2t_{\perp}\cos(k_{\perp}d_{\perp}). \quad (11)$$

Here d_{\perp} is the chain spacing. A small $t_{\perp} \ll t_{\parallel}$ will introduce a warping of the Fermi surface. For considerable interchain coupling, $t_{\perp} \sim t_{\parallel}$, the Fermi surface will be closed, cylindrical.

From Eqs. (8) and (11) a tensorial $\bar{\epsilon}$ may be calculated. By symmetry the principal axes are the chosen directions, along which Drude behaviour is found with plasma frequencies that in the general case must be calculated numerically. As an example to be used later we show in Fig. 4 the normalized plasma frequencies for a quarter-filled band. The crossover from open to closed Fermi surface is indicated.

Electron-phonon coupling

In molecular crystals there are many possible sources to the electron-phonon coupling. Usually, the most important are considered to be: (1) Modulation of t by external modes (acoustic translational and rotational modes), and (2) modulation of E_0 by internal modes (molecular vibrations). A good discussion, including estimates for TTF-TCNQ, has been given by Conwell (1980). The second source to the electron-phonon interaction is of central interest to analysis of IR spectra, since the molecular vibration frequencies span the IR range ($100\text{-}3000\text{ cm}^{-1}$).

The origin of the electron-molecular vibration (emv) coupling is easily understood. The conduction electron orbital energy is in general a func-

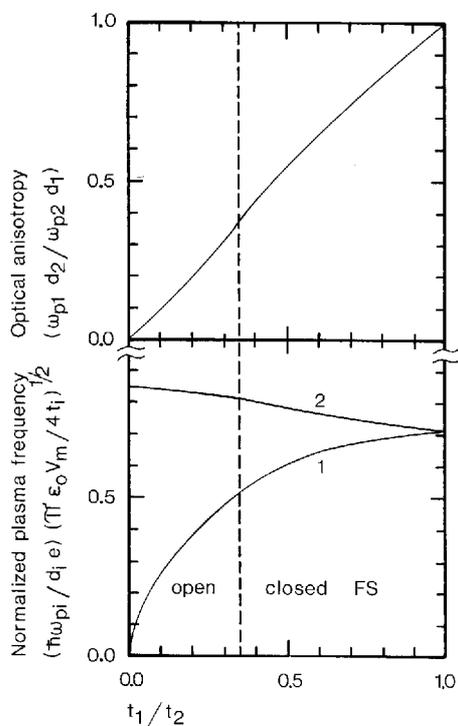
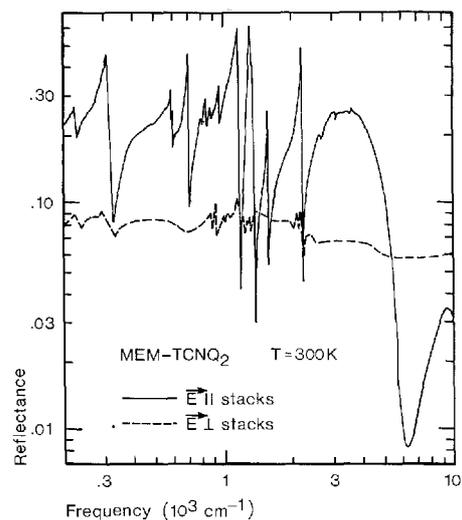


Fig. 4. Anisotropic plasma behaviour in 2D orthorhombic and quarter-filled tight-binding model. The dashed line indicates the cross-over from open (left) to closed (right) Fermi surface. The bottom graphs show normalized plasma frequencies.

tion of the exact atomic configuration in the molecule. Hence the molecular vibrations will modulate the one-electron energies, giving rise to the emv coupling. For a non-degenerate level only the totally symmetric A_g -modes couple linearly to the electrons (Duke et al., 1975). The A_g -modes are Raman active, but IR inactive. However, when charge can move to and from the molecule as in the solid state or in molecular complexes in solution, the modes may, through the emv coupling, borrow oscillator strength from the conduction electrons and give rise to spectacular effects in the IR (see, for example, Bozio and Pecile (1980) and references therein). If the electronic structure is sufficiently simple, information on coupling constants may be extracted from the spectra. One example is the material MEM-TCNQ₂ at 300K, where the TCNQ molecules are organized in stacks as quasi-isolated dimers, with one electron per dimer. In Fig. 5 is shown the IR spectra. Note the strong features in the chain axis spectrum, which has been used to obtain the emv coupling constants for TCNQ (Rice et al., 1980). The coupling constants may also be calculated theoretically and reliable values are now known for several molecules.

Fig. 5. Polarized reflectance of MEM-TCNQ₂ at 300K. Notice the logarithmic scales. (Rice et al., 1980).



In the organic conductor with uniform stacks, the emv coupling should not have direct resonant effects on the optical spectra. Resonances/antiresonances are only activated if the symmetry is broken, either by structural distortion or by the presence of charge-density waves (CDWs) in the system.

Assuming unbroken symmetry, the question arises how electron-phonon interactions influence the IR behaviour. The main deviations from the Drude spectrum are found in the pure metal at low temperature, where the electron-phonon coupling does not contribute to the dc-resistivity. Then the dc-relaxation rate, γ_{dc} , is determined by residual impurities and defects, and a near δ -function contribution to $\sigma(\omega)$ is obtained. As ω is increased and reaches frequencies of current-degrading phonons, photon absorption assisted by phonon emission becomes possible and a threshold in $\sigma(\omega)$ is anticipated (Holstein (1954, 1964); see also Allen (1971)). At frequencies well above all important phonon frequencies, the relaxation rate $\gamma_{opt} \rightarrow \pi\lambda_{tr}\langle\omega_{ph}\rangle$, where λ_{tr} is an appropriate dimensionless electron-phonon coupling constant and $\langle\omega_{ph}\rangle$ is an average phonon frequency. In terms of oscillator strength, a fraction $\sim\lambda_{tr}/(1+\lambda_{tr})$ of the low frequency Drude contribution appears as IR absorption with a high frequency Drude tail determined by phonon emission processes. However, the sum rule, Eq. (10), is still expected to be obeyed.

At high temperatures $k_B T \gg \hbar\langle\omega_{ph}\rangle$, and both phonon absorption

and emission contributes to γ so $\gamma_{dc} \approx \gamma_{opt}$ and a single Drude absorption is found. We note that in the molecular metals the high temperature limit is never reached with respect to the majority of the vibrational modes: Thus Holstein thresholds may be encountered, even at room temperature.

Electron-electron interactions

The electron-electron Coulomb interaction is more difficult to deal with. It is commonly assumed that it can be treated as in ordinary metals. Thus the long range interaction part is frozen in the zero point motion of the plasmons, while the short range part gives rise to quasiparticles with screened interactions (see, for example, Pines (1964)). The screening efficiency was recently discussed by Mazumdar and Bloch (1983). Their conclusions may be described as follows: Only on-site and nearest neighbour interactions need to be considered. Thus the proper model Hamiltonian is of the extended Hubbard type (Hubbard, 1978):

$$H = \sum_{i,\sigma} E_{\sigma} n_{i,\sigma} - \sum_{i,\sigma} t (c_{i,\sigma}^{\dagger} c_{i+1,\sigma} + c_{i+1,\sigma}^{\dagger} c_{i,\sigma}) + \sum_i U n_{i\uparrow} n_{i\downarrow} + \sum_{i,\sigma,\sigma'} V n_{i,\sigma} n_{i+1,\sigma'} \quad (12)$$

U is the extra electrostatic energy associated with two carriers on the same site, and V is the corresponding energy for two carriers on neighbour sites. Both U and V are to be considered effective, screened values, which depend strongly on band filling (ρ). For $\rho = 0.5$ and 1.0 the screening is found to be inefficient ($U > 4t$). In the intermediate range, screening may be quite efficient ($U \ll 4t$), mostly so for $\rho \approx 0.7-0.8$. $\rho = 1$ (half-filled band) constitutes a special case, since for any finite U , there is a gap in the excitation spectrum (Lieb and Wu, 1968). Thus the 1D half-filled band is an insulator of the Mott type. For $\rho < 1$ there is no clear evidence for a gap in the energy spectrum of Eq. (12).

We now turn to the IR properties of a system described by the Hubbard model. First we note the existence of a modified partial sum rule for the intraband absorption (Maldague, 1977):

$$\int_{\text{intraband}} \sigma(\omega) d\omega = -(\pi e^2 d^2 / 2\epsilon_0) \langle H_t \rangle \quad (13)$$

$\langle H_t \rangle$ is the ground state expectation value of the transfer terms of H . Since H_t determines the ground state in the absence of correlation effects, it is evident that finite correlation reduces the oscillator strength below that of Eq. (10).

The next issue is the distribution of oscillator strength. Again $\rho = 1$ is a special case. Even in the weak coupling limit, $U \ll 4t$, optical absorption may occur via Umklapp electron-electron scattering. In the strong coupling limit, $U \gg 4t$ ($V=0$) Lyo and Gallinar (1977) find a symmetric absorption extending from $U-4t$ to $U+4t$. The oscillator strength is proportional to t^2/U . A finite nearest neighbour interaction induces an asymmetry in the absorption (Lyo, 1978). Then there is a strong onset of absorption at $U-4t$ and a tail extending to $U+4t$. Physically the absorption band corresponds to charge transfer transitions into states with doubly occupied sites.

For $\rho < 1$, the main question is whether the dominant part of the oscillator strength goes into charge transfer bands of the correlation type ($\hbar\omega \sim U, V$) or whether it appears in a low frequency Drude-like absorption. The answer again depends on the band-filling. However, there is strong evidence from various approximative calculations (Maldague, 1977) and finite chain calculations that for $0.5 < \rho < 0.6$ less than 10% of the oscillator strength appears in correlation bands.

Finally, it is of some interest to describe a specific limit of Eq. (12): $V = 0$ and $U \rightarrow \infty$ (double occupancy of sites excluded). In this case the translational and the spin degrees of freedom are decoupled (Sokoloff, 1970). The spin susceptibility is Curie-like and the electrons behave otherwise as a system of non-interacting spinless fermions described by the usual tight-binding model of Eq. (5,6). Because of the spinlessness the density of states is halved and the band is now filled to $k = \pm\pi\rho/d$. The optical properties are Drude-like, as described above, and the relative reduction of oscillator strength is obtained from Eq. (9):

$$\int_{U \rightarrow \infty} \sigma(\omega) d\omega \Big/ \int_{U=0} \sigma(\omega) d\omega = \cos(\pi\rho/2). \quad (14)$$

It is noted that with respect to oscillator strength, the dense electron system is more sensitive to the high U limit than the less dense system.

Low-dimensional instabilities

The linear chain structure of the molecular conductors has important consequences for the occurrence and behaviour of instabilities. As discussed below the quasi-one-dimensionality leads to phase transitions not known in isotropic systems. At the same time thermodynamic fluctuations tend to suppress the phase transition temperatures. It is well-known that long range order cannot persist in a 1D system for $T > 0$. Thus usually short range order develops below some scale temperature,

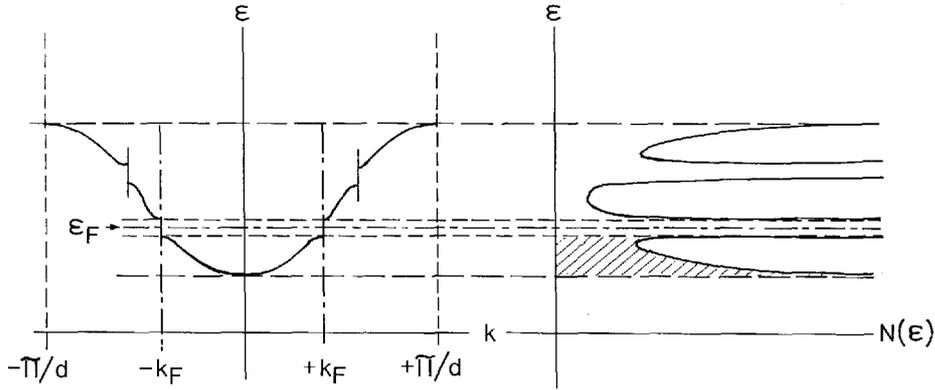


Fig. 6. Effect of $2k_F$ -distortion on cosine band and density of states.

T_{MF} . As temperature is lowered, the on-chain correlation length grows and since the real systems consist of arrays of chains any finite interchain coupling will eventually induce long range 3D order at a temperature, T_C , which defines a three-dimensional phase transition. In the range $T_C < T < T_{MF}$, the fluctuations may have important impact on the physical properties.

The most famous of the 1D instabilities is that of the linear chain metal. We shall here repeat the argument by Peierls (1955) with reference to Fig. 6. The figure shows the band structure and density of states of Eq. (6) with the inclusion of a weak $2k_F$ -potential which spans the Fermi sea. The crucial point is that this potential opens a gap at the Fermi level, ϵ_F , and thus lowers the energies of those electrons closest to ϵ_F . Due to the 1D divergent density of states near a gap this energy gain will always outweigh the cost of creating the $2k_F$ -potential. The latter arises from a CDW/periodic lattice distortion mediated by the electron-phonon coupling. For a half-filled band a detailed calculation gives the following zero temperature gap (Rice and Strässler, 1973):

$$E_g = 2\Delta = 2\epsilon_F / \sinh(1/\lambda - \ln 2) \approx 8\epsilon_F \exp(-1/\lambda). \quad (15)$$

The last expression applies for $\lambda \ll 1$. The dimensionless electron-phonon coupling constant, λ , is related to the bare coupling constant, g , the density of states in the metallic state, $N(\epsilon_F)$, and the unperturbed phonon frequency, ω_{2k_F} , by

$$\lambda = g^2 N(\epsilon_F) / \hbar \omega_{2k_F}. \quad (16)$$

The temperature dependence of the gap is BCS-like (Bardeen et al., 1957). Thus the mean field scale temperature, where the gap vanishes is given by

$$k_B T_{MF} = 2\Delta/3.52. \quad (17)$$

The actual phase transition temperature, T_C , may be fluctuation suppressed below T_{MF} .

There are two kinds of elementary excitations associated with the Peierls state: (1) Electron-hole excitations across the gap and (2) phase and amplitude oscillations of the $2k_F$ -CDW. The phase oscillations have a dipole moment and are thus IR active. Fröhlich (1954) noted that in situations where the energy of the CDW is independent of the phase, the CDW is free to move through the crystal and carry current in a way reminiscent of superconductivity. In real systems this may happen in incommensurate cases (where λ_{2k_F} does not match the lattice) for $T \geq T_C$, where, however, a finite correlation length will limit the conductivity. For $T < T_C$ the CDW is pinned and appears as a far IR absorption.

Rice and coworkers (Rice et al. (1975), Rice (1978)) have emphasized the role of simultaneous involvement of many phonons. They find that for the molecular conductors, it is to be expected that much of the Peierls gap is due to the sizeable number of intramolecular modes coupled to the electrons, while the dynamic properties, e.g. effective mass, of the CDW is dominated by the low frequency external modes. The CDW is a complicated superposition of contributions from each mode. In the infrared spectrum phase oscillations may be identified in the vicinity of each unperturbed mode frequency (Rice et al., 1977). This is a good example of IR-activation of the A_g -modes.

A number of other instabilities have close analogies to the Peierls instability. The Overhauser instability (Overhauser, 1960) may occur when the electron-phonon interaction is comparatively weak and a finite short range electron-electron interaction is present. Then a $2k_F$ spin-density wave will act on the electrons as an effective $2k_F$ -potential through the modulation of the exchange interaction (Slater, 1951). The energetics are then similar to what is described above, but the effective coupling constant rather is $N(\epsilon_F)\tilde{U}$, \tilde{U} being an appropriate Coulomb interaction.

Another important modification of the Peierls instability happens in the limit of strong electron-electron interaction, where we may let $U \rightarrow \infty$. As stated above, the translational degrees of freedom are described in terms of non-interacting spinless fermions in a tight-binding band filled

to $\pm 2k_F$. The fermion-phonon interaction is of the same form as that entering the theory for the Peierls instability, thus a $4k_F$ -CDW Peierls type of instability may occur (Bernasconi et al., 1975). The behaviour is completely analogous if the differences in wavenumber, in band-filling, and in dimensionless coupling constant are taken into account.

If the electrons are locked into such a $4k_F$ -CDW or are localized by direct Coulomb interaction (Mott insulator) the magnetic subsystem is susceptible to yet another $2k_F$ -instability, the spin-Peierls instability (Chesnut, 1966), which is a magnetic analogue. In this case the coupling between the antiferromagnetic exchange integral, J , and the phonons induces a $2k_F$ modulation of J , which quenches the magnetic susceptibility.

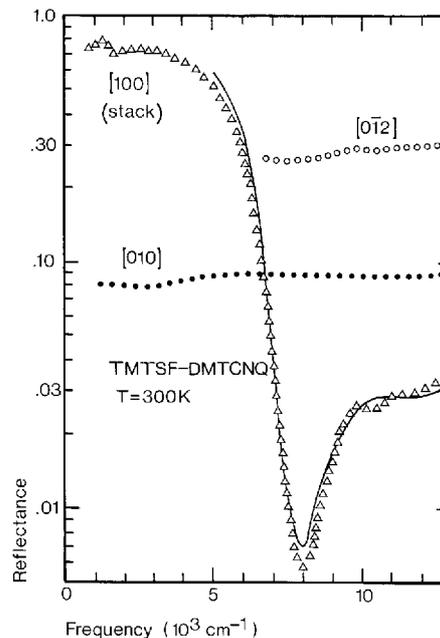
After this short discussion we stress that for the instabilities to occur, the 1D nature is crucial. The Peierls type instability depends on the ability of a single wavevector to nest the entire Fermi surface. The nesting efficiency will in general deteriorate as interchain coupling increases (see, for example, Horowitz et al. (1975)), but even 3D materials may undergo density wave formation if segments of the Fermi surface allow nesting.

As a last remark it is noted that also superconductivity may occur in 1D systems. However, it is not clear at present how one may distinguish between a quasi-1D superconductor and a highly anisotropic 3D superconductor.

Plasma edges and band structure

Measurements of the polarized reflectance in the vicinity of the plasma edge have frequently been used to obtain estimates for the transfer integrals. The method consists of fitting a reflectance model based on the Drude expression of Eq. (7) to the data in a limited range and then use Eq. (8) in conjunction with information on carrier density and crystallographic data to derive values for t . Such an approach has often been criticized (see, for example, Williams and Bloch (1974)) because the molecular conductors show strong deviations from Drude behaviour at lower frequencies (cfr. next section). The aim of the present section is to use experimental data in arguing for the validity of the Drude analysis. The underlying idea is that the plasma edge is a signature for existence of plasmons. Since these are long wavelength oscillations in the charge density their frequency is intuitively expected to be insensitive to the

Fig. 7. Polarized reflectance of TMTSF-DMTCNQ at 300K. Notice the logarithmic reflectance scale. The solid line is a fit of the Drude model to the chain axis spectrum. An extra oscillator has been added to the model to account for the absorption band near 10000 cm^{-1} .



details in the short range interactions, which give rise to the non-Drude features observed in the infrared. Thus the Drude expression should be viewed as a convenient mathematical tool in estimating the plasmon frequency, Ω_p , and the background dielectric constant, ϵ_c . Then the transfer integral is calculated from the unscreened plasma frequency, $\omega_p \approx \Omega_p \sqrt{\epsilon_c}$.

During the argumentation we shall further attempt a decomposition of average transfer integrals in double-stack conductors and show examples of materials with 2D character.

Chain axis

As a typical example we present in Fig. 7 the polarized room temperature reflectance of TMTSF-DMTCNQ with a model fit to the chain axis spectrum. The most striking feature is the well-developed plasma edge observed along the stacks. Virtually no dispersion is seen in the other directions. The infrared spectra are indeed a remarkable manifestation of the linear chain structure.

The plasma edge is quite sharp with a drop in reflectance of more than two orders of magnitude. This hints to the existence of rather well-defined plasmons, i.e. $\epsilon_1 = 0$ and $\epsilon_2 \ll 1$ are simultaneously obeyed.

Plasmons have so far been observed directly in TTF-TCNQ only (Ritsko et al., 1975). In that case the plasmon frequency is in complete agreement with the optical properties (Tanner et al., 1976).

The solid line is a fit of a model based on the Drude expression, Eq. (7), with an extra oscillator added. The latter takes into account the absorption band near 10000 cm^{-1} , which is observed in all TCNQ based materials and which is presumably intramolecular in origin (Tanaka et al., 1978).

A close inspection of the quality of the fit reveals a systematic deviation from the Drude model in the edge region. The data points fall below the model both at low frequencies (entering the IR) and near the minimum. The model could be improved by allowing the relaxation rate, γ_{opt} , to decrease slowly with frequency in the near infrared range. This observation, which is made in many of the organic conductors leads us to shortly discuss the origin of optical absorption in the near IR. Usually the absorption is attributed to phonon assisted electron-hole excitations within the one-electron tight-binding band. However, the bandwidth, $4t$, is found to be of order 0.7-1.0 eV corresponding to frequencies of $6000\text{-}8000\text{ cm}^{-1}$. But as the frequency exceeds the bandwidth many phonon assisted processes become impossible by energy conservation. For $\hbar\omega > 4t$ only processes involving emission of high energy phonons are conceivable. Thus such absorption mechanisms in a natural way account for the observed frequency dependence of γ_{opt} . This discussion viewed together with the remarks made on the Holstein mechanism previously, strongly suggests that the near IR relaxation rate has little to do with the low frequency scattering, which governs the static conductivity. Indeed, the temperature dependence of γ_{opt} is much weaker than that of the dc-conductivity in most cases (see, for example, Bright et al. (1974)).

In Table III is presented the results of Drude analysis and transfer integral calculation for a number of materials, in all cases for the stacking direction, using Eq. (9). For the double-stack conductors the value given is the average of donor and acceptor bandwidths. For the single-stack conductors the slight dimerisation is neglected.

Since the table contains some results for single- and double-stack conductors with the same molecule, it should be possible to decompose the bandwidth of the double-stack conductors into donor and acceptor contributions adopting certain assumptions. For this purpose interplanar distances (ipd) for the molecules in the stacks have been included. Our principal assumptions are, (1) that the bandwidth depends only on the

Table III. Chain axis Drude parameters and bandwidths. Also given are cation and anion interplanar distances.

Material	T K	ϵ_c	γ cm^{-1}	ω_p cm^{-1}	$4\bar{t}$ eV	cation ipd \AA	anion ipd \AA
TTF-TCNQ	300	3.27	1430	11400	0.61	3.48	3.17
TSF-TCNQ ^a	300	—	—	12200	0.69	3.52	3.21
TMTSF-TCNQ ^a	300	—	—	12000	0.89	3.59	3.26
TMTTF-DMTCNQ ^{a,b}	300	—	—	10700	0.77	—	—
TMTSF-DMTCNQ	300	2.91	1180	11200	0.88	3.64	3.31
HMTTF-TCNQ	300	3.15	1260	12400	0.78	3.57	3.23
HMTSF-TCNQ	300	3.30	1030	14200	1.03	3.6	3.2
DBTTF-TCNQCl ₂	300	2.51	2000	7500	0.41	3.51	3.41
HMTSF-TNAP	300	2.95	1100	11300	0.84	3.58	3.28
TMTTF ₂ PF ₆	300	2.50	1380	8900	0.80	3.58	—
TMTSF ₂ AsF ₆	300	2.56	1230	9900	1.00	3.63	—
—	30	2.55	1160	10500	1.11	—	—

^a estimate from edge shift

^b q is assumed equal to 0.5 (as in the isostructural TMTSF-DMTCNQ)

molecular species and ipd, and (2) that t increases $\sim 5\%$ for a decrease in ipd of 1%. The first assumption is based on the quasi-1D nature of the materials and on the very similar overlap patterns observed for the same molecule in different materials (see references of Table II). The second assumption is based on molecular orbital calculations (Herman, 1977) and on the change in t observed on cooling to low temperatures. This change usually amounts to $+10\%$ (Table III), while the stack-contraction is about 2% (Schultz et al., 1976).

The resulting decomposition is given in Table IV. There is a rather large spread in bandwidths from 0.4 to 1.3 eV. Substituting Se for S in TTF-based molecules always results in increases in t as expected from the increased overlap. It is noteworthy that methylation of TTF-based molecules leads to strong increases in bandwidths (Jacobsen et al., 1983). Although we cannot go into a detailed discussion of the results in context of other physical properties the general trends are clearly consistent with independent knowledge. We give four specific examples: (1) The thermoelectric power of TMTSF single stack materials yields a bandwidth slightly larger than 1 eV (Mortensen, 1982). (2) HMTSF-based metals have exceptionally high conductivities (cfr. Table II). (3) The thermoelectric power of TTF-TCNQ is large and negative while it is close to zero for TSF-TCNQ (Chaikin et al., 1976). (4) The conductivity of

TMTSF-TCNQ is significantly higher than that of TSF-TCNQ (Table II).

Another argument for the validity of the present approach is the agreement in values for the DMTCNQ-bandwidths. The number obtained for TMTTF-DMTCNQ is based on analysis of data for the single-stack compound TMTTF₂PF₆, which is a semiconductor with strong non-Drude behaviour in the IR (Jacobsen et al., 1983), while the number for the selenium analogue is based on data for an excellent metal, TMTSF₂AsF₆, which is close to Drude behaviour in the IR.

A final direct demonstration of the main point i.e. the simple connection between one-electron bandwidth and plasmon frequency will be given next while discussing 2D compounds.

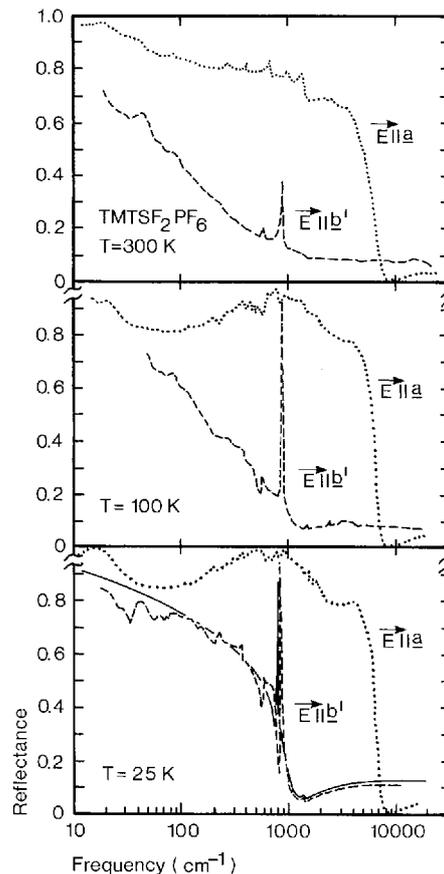
Interchain contact

There are at present two groups of organic conductors, which show considerable interchain coupling. They are both of single-stack type with donor molecules, $\rho = 0.5$, and inorganic counterions. The basic structural features are also the same: Sheets of molecular stacks interchanging with layers of counterions. The important interchain contact is *in* the sheets so that the materials are effectively two-dimensional. Interestingly these materials among them count the only representatives of organic superconductivity so far.

Table IV. Single-stack bandwidths. T = 300K

Donor		Acceptor	
Stack	4t eV	Stack	4t eV
TTF-(TCNQ)	0.41	(TTF)-TCNQ	0.81
TSF-(TCNQ)	0.62	(TSF)-TCNQ	0.76
TMTSF-(TCNQ)	1.06	(TMTSF)-TCNQ	0.70
TMTTF-(DMTCNQ)	0.80	(TMTTF)-DMTCNQ	0.74
TMTSF-(DMTCNQ)	0.93	(TMTSF)-DMTCNQ	0.70
HMTTF-(TCNQ)	0.80	(HMTTF)-TCNQ	0.76
HMTSF-(TCNQ)	1.3	(HMTSF)-TCNQ	0.76
HMTSF-(TNAP)	1.3	(HMTSF)-TNAP	0.4

Fig. 8. Polarized reflectance of $\text{TMTSF}_2\text{PF}_6$ at 300K, 100K, and 25K. Notice the logarithmic frequency scale. The chain axis is along a . The solid line is a Drude fit to the low temperature spectrum with polarization perpendicular to the chains. One extra oscillator has been added to the model to account for the sharp vibrational line near 800 cm^{-1} . (Jacobsen et al. 1981).



Their actual anisotropy may be deduced from reflectance studies. As an example from the TMTSF_2X -family we present in Fig. 8 the polarized reflectance of $\text{TMTSF}_2\text{PF}_6$ at three temperatures: 300K, 100K and 25K (Jacobsen et al., 1981). It is evident that a reasonably well-defined plasma edge appears in the b' -direction at low temperature. The b' -direction is perpendicular to the stacks in the sheets of TMTSF -molecules. Most of the sharp lines superimposed on the edge arise from normal IR active modes in PF_6^- . The transverse reflectance edge appears at a frequency about ten times lower than that of the stacking axis edge. A Drude analysis may be performed (solid line) and the ratio $(\omega_{pb'}/\omega_{pa'})$ (a'/b') which refers to the model of Fig. 4 is 0.09. Calculation yields $t_{b'} \approx 22\text{ meV}$, about 10 times smaller than t_a . Although $t_{b'}$ is significant as compared to $k_B T$ for all temperatures of interest the Fermi surface must clearly be open. Table V gives Drude parameters and transfer integrals

Table V. Drude parameters, transfer integrals, and b-axis lattice constants in $(\text{TMTSF})_2\text{X}$, $\vec{E} \parallel b'$

X	ϵ_c	ω_p cm^{-1}	γ cm^{-1}	$t_{b'}$ meV	b \AA
ClO_4^-	3.5	2020	250	24	7.678
SbF_6^-	3.5	1510	300	18	7.728
AsF_6^-	3.5	1670	350	20	7.711
PF_6^-	3.5	1830	500	22	7.711

for several members of the TMTSF_2X family (Jacobsen et al., 1983). It should be noted that there is a reasonable correlation between $t_{b'}$ and the b-axis lattice constant. The largest $t_{b'}$ is found in the densely packed $\text{TMTSF}_2\text{ClO}_4$, the only ambient pressure superconductor in the series. This gives some support to the simple band picture of Yamaji (1985) which predicts a critical value of $t_{b'} \approx 25$ meV. The values are also in good agreement with results of band structure calculations (Grant, 1983).

The other materials group of interest is based on BEDT-TTF (or ET). The first reported ambient pressure superconductor in this family is triclinic (β -phase) ET_2I_3 (Yagubskii et al., 1984). Fig. 9 shows the polarized reflectance of this material in the a - b plane (Jacobsen et al., 1985). Again this plane corresponds to the sheets of molecular stacks. Interestingly, the highest plasma frequency is associated with the direction perpendicular to the chains, while the chain direction shows much weaker metallic character in the infrared properties. Although unusual, such behaviour is consistent with the molecular arrangement in the crystal (Kaminskii et al., 1984) and has also been seen optically at room temperature in several materials from the family (Tajima et al. (1984) and Koch et al. (1985)).

In Table VI we list Drude parameters and estimated transfer integrals for ET_2I_3 . Again the results of Fig. 4 have been used. Thus the analysis is based on a simplified orthorhombic model and the transfer integrals given are weighted averages of several transfer integrals in the correct triclinic model. The rather small values of t are consistent with a low room temperature conductivity ($\sim 30 \text{ Scm}^{-1}$, isotropic in the a - b -plane) and the near isotropy is confirmed by thermopower measurements of Mortensen et al. (1985). The validity of the plasma edge Drude analysis

Table VI. Drude parameters and transfer integrals in $\beta\text{-ET}_2\text{I}_3$.

|| Denotes the chain axis, \perp the direction perpendicular to the chains in the a - b plane.

axis	T K	ϵ_c	ω_p cm^{-1}	γ cm^{-1}	$\langle t \rangle$ eV
	40	4.0	5740	2270	0.10
\perp	40	3.8	9600	540	0.13
\perp	300	3.8	9100	1360	0.12

is again confirmed. It is particularly interesting to compare the 40K and 300K analysis of the data for the perpendicular direction. It is obvious from Fig. 9 and even more so from the dielectric function reproduced in Fig. 10 that the 300K spectrum is non-Drude in the IR while the 40K spectrum is near-Drude in the entire measured range. Nevertheless, the predicted plasma frequencies are different only by an amount expected

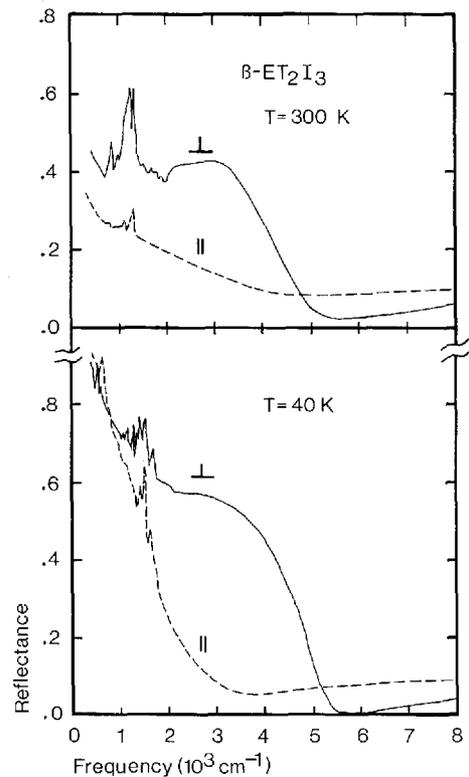


Fig. 9. Polarized reflectance of $\beta\text{-ET}_2\text{I}_3$ at 300K and 40K. The stacking axis is designated ||, while \perp refers to a direction perpendicular to the stacks in the a - b plane. (Jacobsen et al., 1985).

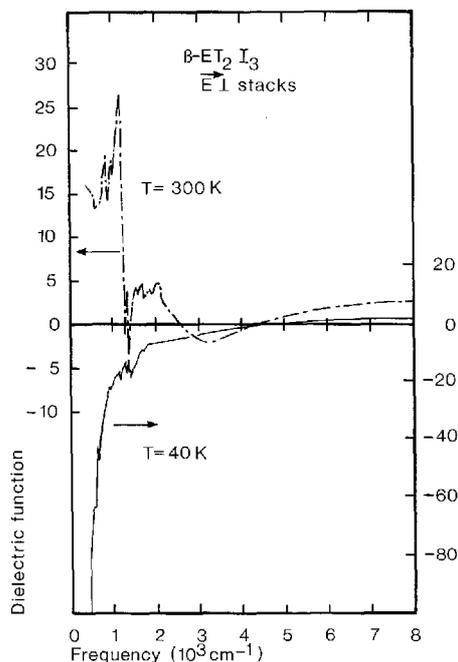


Fig. 10. Real part of $\epsilon(\omega)$ for β - ET_2I_3 perpendicular to the chains, for $T=300\text{K}$ and 40K . Notice the different scales for the two curves. (Jacobsen et al., 1985).

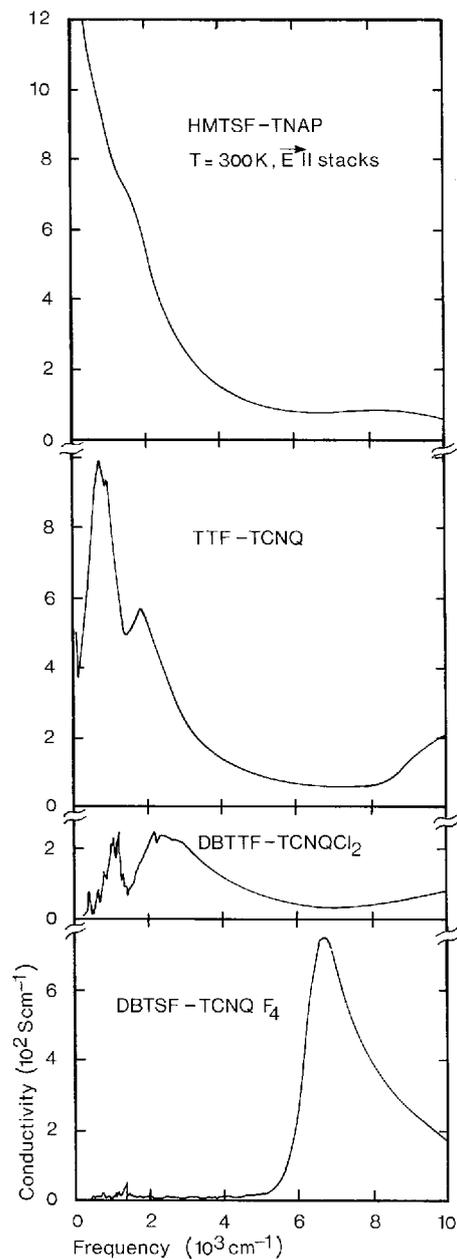
from a thermal contraction induced change in t_1 (Table VI). This is also evident from Fig. 10 by noting that the zero-crossing in $\epsilon(\omega)$ near 5000cm^{-1} only shifts slightly.

Hence our principal conclusion of this section is that careful analysis of reflectance data in the plasmon region yields reliable estimates for transfer integrals. These estimates will be used subsequently while discussing the IR spectra.

Size and distribution of infrared oscillator strength

After having discussed the plasma edge range we now turn to the actual infrared excitation spectra of organic conductors. The aim is to understand the deviations from Drude behaviour in terms of the short range electron-electron and electron-phonon interactions. In Fig. 11 is shown 300K-spectra of 4 double-stack conductors with widely different physical properties (Jacobsen et al., 1984). HMTSF-TNAP, TTF-TCNQ and DBTTF-TCNQCl₂ are incommensurate conductors with slightly more than quarter-filled one-electron bands, while DBTSF-TCNQF₄ has half-

Fig. 11. Frequency dependent conductivities of 4 doublestack conductors at $T=300\text{K}$, along the stacks. (Jacobsen et al., 1984).



filled bands. All materials are expected to be metals in the absence of Coulomb interactions and at 300K there is no evidence for the $2k_F$ -distortion, which might destroy the metallic state by pure structural effects.

DBTSF-TCNQF₄ is known to be a Mott insulator (Lerstrup et al., 1983), and as we shall shortly discuss, available theory allows a rather unambiguous interpretation of the spectrum.

HMTSF-TNAP is a high conductivity, wide band metal, and the IR behaviour is indeed rather close to Drude behaviour with a monotonic $\sigma(\omega)$. TTF-TCNQ is also a metal but of intermediate bandwidth and with distinct deviations from the Drude spectrum. The oscillator strength is shifted away from zero frequency and fine structure is seen near molecular vibration frequencies. These features are even more pronounced in the narrow band material DBTTF-TCNQCl₂, which is a magnetic semiconductor. It is obvious that these room temperature spectra can not be rationalized within traditional 3D models like the Holstein absorption picture.

In the following we first derive parameters for the Mott insulator case. Next we review the physical properties of DBTTF-TCNQCl₂ and argue for using the simple $U \rightarrow \infty$ model for this material. Then the size of the IR oscillator strength for a number of organic conductors is compared with expectations based on bandwidth estimates. It is suggested that Coulomb correlations modify the wavefunctions appreciably in all materials. However, depending on band-filling and dimensionality the effect on other physical properties differs widely. Finally, we discuss the temperature dependent IR properties of TTF-TCNQ with special emphasis on understanding the separate roles of the two chains.

The Mott insulator ($Q = 1$)

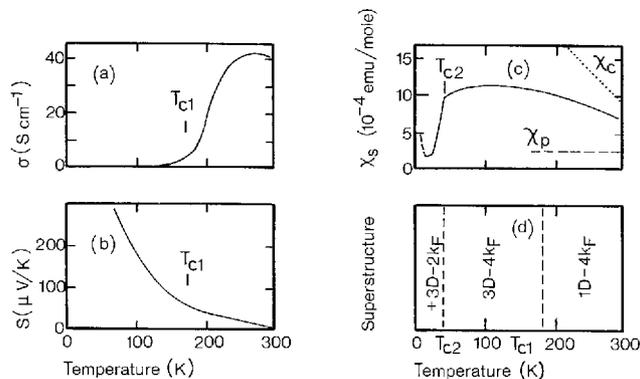
The frequency dependent conductivity of DBTSF-TCNQF₄ resembles surprisingly well the prediction by Lyo (1978) for the extended Hubbard model with half-filled one-electron bands. The absorption is dominated by a charge transfer band corresponding to the creation of a doubly occupied site. A finite near neighbour interaction, V , produces a rather strong onset of absorption at $U-4t$ as observed near $5500 \text{ cm}^{-1} \approx 0.7 \text{ eV}$ in the present case. Additional information may be obtained from the size of the oscillator strength. It is assumed that it follows that of the isolated Hubbard dimer (Rice, 1979):

$$\int \sigma(\omega) d\omega / \int_{(U=0)} \sigma(\omega) d\omega = (1 + (U/4t)^2)^{-1/2}, \quad (18)$$

(correct for $U/4t \ll 1$ and $\gg 1$ (Lyo, 1978)).

Then we find $U \approx 1.4 \text{ eV}$ and $4t \approx 0.7 \text{ eV}$. Such a value for the bandwidth appears reasonable when compared to the results of Table III

Fig. 12. Basic physical properties of DBTTF-TCNQCl₂. See the discussion in the text.



for similarly composed materials (e.g. TSF-TCNQ) but an independent estimate is not available. $U/4t \approx 2$ for the half-filled band case is in accord with theoretical expectations (Mazumdar and Bloch, 1983).

As a side-remark it is noted that the A_g -vibrational modes are not activated in the Mott insulator. Charge localization does not suffice: symmetry breaking is a necessary prerequisite for IR activation.

The infinite U model ($\rho < 1$)

DBTTF-TCNQCl₂ is a good example of a non-metallic, but incommensurate ($\rho = 0.56$) and near quarter-filled organic conductor. The physical properties are known in rather great detail and are summarized in Fig. 12. The dc conductivity and the thermoelectric power are consistent with the existence of a one-electron gap of order 200-250 meV ($1500-2000$ cm⁻¹). The activated behaviour is cleanest below $T_{c1} = 180$ K where a slight anomaly is observed. The spin susceptibility, χ_s , is high, about three times the expected non-enhanced Pauli value (χ_p) at 300 K, and remains high to $T_{c2} = 36$ K, below which temperature it vanishes rapidly. Note that χ_s (300 K) is about $\frac{2}{3}$ of the Curie value (χ_c) for the appropriate carrier density. Finally, diffuse X-ray scattering has shown the existence of rather large amplitude $4k_F$ -CDWs. Down to T_{c1} there are only weak interchain correlations. T_{c1} is found to be a 3D ordering temperature, while T_{c2} involves the onset of 3D- $2k_F$ scattering.

The implications of these observations are obvious: Enhanced magnetic susceptibility and occurrence of $4k_F$ -CDWs may both be taken as evidence for important electron-electron correlations. The two transition temperatures also indicate a substantial decoupling of the spin and translational degrees of freedom. The overall behaviour is common to a

larger group of low conductivity materials and the physical picture is the following: The carriers are fairly localized even at high temperature due to the Coulomb repulsion between them. The spins experience a weak antiferromagnetic exchange coupling and through interaction with the lattice undergo a spin-Peierls transition at low temperature. With this discussion we have established that DBTTF-TCNQCl₂ is a strongly correlated, incommensurate organic conductor with a sizeable energy gap in the excitation spectrum.

In Fig. 13(a) $\sigma(\omega)$ is shown at 100K. The main difference from the 300K spectrum is a sharpening of the vibrational structure below 1500 cm⁻¹. In both spectra there is a broad maximum around 2000 cm⁻¹, which may be interpreted as arising from a gap in the electronic spectrum in accordance with the low frequency transport properties. This gap, the intensity of the vibrational structure, the presence of 4k_F-CDWs, and the apparent decoupling of magnetism and electronic behaviour leads us to suggest that DBTTF-TCNQCl₂ is subject to an »infinite«-U Peierls instability partly stabilized by the emv coupling. A zero temperature model for $\sigma(\omega)$ of the ordinary 2k_F (U=0) Peierls semiconductor has been given by Rice (1978). Because of the formal analogy with this case the theory can be immediately applied, when proper account for the spinlessness of the fermions is taken. As stated earlier the difference concerns the band-filling and the density of states. Rice gives the following expression for $\bar{\epsilon}(\omega)$:

$$\bar{\epsilon}(\omega) = \epsilon_c + (\omega_f/\omega)^2 [f(\omega/2\Delta) - 1 - (\omega/2\Delta)^2 f^2(\omega/2\Delta) \lambda D_\varphi(\omega)]. \quad (19)$$

ω_f is a measure for the oscillator strength so that

$$\int_0^\infty \sigma(\omega) d\omega = (\pi/2) \epsilon_0 \omega_f^2. \quad (20)$$

2 Δ is the Peierls gap and λ is the total electron-phonon coupling constant given by

$$\lambda = \sum_n \lambda_n = \sum_n g_n^2 N(\epsilon_F) / \hbar \omega_n, \quad (21)$$

where n is the mode number.

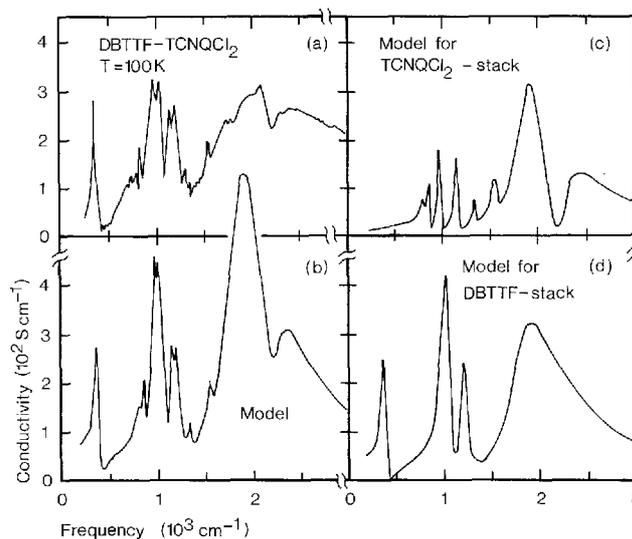
Furthermore, the function, $f(x)$, is defined by

$$f(x) = [\pi i + \ln(1-S)/(1+S)] / 2Sx^2, \quad S = (1-x^{-2})^{1/2} \quad (22)$$

A small electronic damping may be introduced by the substitution $x^2 \rightarrow x(x+i\delta)$ in these expressions. Finally, the phase phonon propagator

$$D_\varphi(\omega) = [D_0^{-1}(\omega) + 1 + \lambda(\omega/2\Delta)^2 f(\omega/2\Delta) + b]^{-1} \quad (23)$$

Fig. 13. $4k_F$ -CDW model for $\sigma(\omega)$ of DBTTF-TCNQCl₂ at 100K. (a) data, (b) total $\sigma(\omega)$ in model, (c) contribution from acceptor stack, and (d) contribution from donor stack.



where

$$D_o(\omega) = -\sum_n (\lambda_n/\lambda) \omega_n^2 / (\omega_n^2 - \omega^2 - i\omega\gamma_n). \quad (24)$$

γ_n is the natural damping rate of mode n . b is a positive constant, which models the pinning of the CDW; if $b = 0$, the CDW contributes to the dc conductivity.

The physical contents of this model were described previously. The main problem in its application to DBTTF-TCNQCl₂ is the double-stack structure. There are two contributing stacks and we must assume that $\tilde{\epsilon}(\omega)$ is simply a sum of two individual terms. Clues to the decomposition may be obtained by studying single-stack materials like DBTTF(BF₄)_{0.42}. It is rather easy to show that DBTTF dominates the full spectrum and we have obtained a model fit to the data of the quality shown in Fig. 13(b). The individual contributions are pictured in Fig. 13 (c+d). The model parameters are given in Table VII, which also contains independent estimates of the emv coupling constants.

The overall quality of the fit is reasonable, especially with respect to distribution of oscillator strength between the vibrational modes and the single particle contribution. We note that it is possible to introduce only a modest electronic damping in the present model hence the gap structure in the fit is sharper than in the data.

From Table VII it follows that the coupling constants (g_n) obtained are of the correct magnitude. This is a crucial point, since in Eq. (21) we have used the $U \rightarrow \infty$ density of states, which for a quarter-filled band is

Table VII. $4k_F$ -CDW parameters for DBTTF-TCNQCl₂, T = 100K.

	DBTTF		TCNQCl ₂			
$\omega_f(\text{cm}^{-1})$	4500		3600			
$4t(\text{eV})$	0.43		0.27			
$2\Delta(\text{cm}^{-1})$	1700		1900			
δ	0.2		0.2			
λ	0.46		0.60			
b	0.08		0.08			
$\Delta\varepsilon_c$	1.0		1.0			
n	ω_n cm ⁻¹	λ_n	g_n^a cm ⁻¹	ω_n cm ⁻¹	λ_n	g_n^b cm ⁻¹
1	1430	.16	1100 (940)	2253	.04	550(350)
2	1130	.017	320(170)	1573	.01	230(540)
3	473	.11	530(630)	1345	.01	210(500)
4	(40	.17	190)	1200	.04	400(300)
5				1020	.03	340(85)
6				870	.009	160
7				820	.012	180
8				(40	.45	250)

^a values in parentheses are estimates for TTF for corresponding modes (Lipari et al., 1977).

^b values in parentheses are estimates for TCNQ for corresponding modes (Rice et al., 1980).

only 35% of the $U=0$ value. Thus the strength of the vibrational modes constitutes a direct verification of the applicability of the large U model.

A few other points may be emphasized: (1) Assuming mean field behaviour the observed gap position corresponds to a scale temperature $T_{MF} \approx 600\text{K}$. This is consistent with the well-established gap and CDW-structure at 300K and below. (2) In the fitting procedure the estimated gap-values were used to derive the total electron-phonon coupling constant, λ , from the gap equation (Eq.(15)) for each stack. Part of λ is assumed associated with low frequency (external) modes outside the experimental range (given in the last lines of Table VII). However, it is striking that more than 60% of λ for the DBTTF-stack is due to the internal modes. Hence, the CDW is predominantly stabilized by the env

coupling. On the acceptor chain this is less pronounced. (3) The average bandwidth, $4\bar{\epsilon}$, is 0.35 eV from Table VII. This may be compared to $4\bar{\epsilon} = 0.41$ eV from the Drude analysis (Table III). It appears that the oscillator strength is further reduced below what is expected from the large U model. Introduction of a finite nearest neighbour interaction, V , in the theory might account for that.

In conclusion the IR properties of DBT⁺TF-TCNQCl₂ are surprisingly well described by the model for the large U , $4k_F$ -Peierls system. The infrared properties of several singlestack conductors like TMTTF₂PF₆, and MEM-TCNQ₂ at 350K, may also be analysed satisfactory in terms of the same model (C. S. Jacobsen, to be published). Thus it appears that for a large group of organic conductors the electronic properties are reminiscent of those of a system, where double occupancy of sites is effectively excluded.

Sum rules and correlation effects

To investigate whether the infrared oscillator strength can give information on the strength of Coulomb correlations in general we give in Table VIII a comparison of the observed oscillator strength with the predicted in the $U=0$ and $U\rightarrow\infty$ limits for a number of compounds. In particular, the last column gives the ratio, β , between observed reduction of oscillator strength and the reduction associated with the $U\rightarrow\infty$ model. Thus $\beta=0$ corresponds to $U=V=0$ and $\beta=1$ to $U\rightarrow\infty$, $V=0$. It is notable that β is in no case smaller than 0.5. HMTSF-TCNQ and -TNAP, and TMTSF₂ClO₄, which are all excellent metals have β 's close to 0.5. The semiconducting materials have $\beta=0.9-1.3$. TTF-TCNQ which will be discussed in more detail below has $\beta = 0.85$ decreasing to 0.56 at 60K, the temperature of maximum conductivity. This may indicate a connection between the absolute value of σ and the screening efficiency. A similar trend is observed in ET₂I₃(1). However, here $\beta > 1$ in spite of the metallic ground state. The large β is attributed to the narrow (Table VI) and quarter-filled band. It appears that the metallic/superconducting ground state must be related to the 2D-nature of this material which precludes, for example, the $4k_F$ -CDW instability.

The only material, which appears not to match the pattern is TMTSF-DMTCNQ, which has metallic character, but $\beta = 1.1$. We understand such a large β as arising from a Coulomb correlation dominated acceptor stack and a metallic donor stack as previously implied by analysis of the transport properties (Jacobsen et al., 1978).

Thus it largely appears that the β -values of Table VIII are consistent

Table VIII. Infrared oscillator strength and correlation effects. The first column gives the values expected if $U=V=0$ (based on Tables III and VI). The second column corresponds to $U \rightarrow \infty, V=0$ (Eq. 14), while the third gives the measured oscillator strength. β is a relative measure for the strength of the correlation effects (see text).

Material (temp.)	ω_p^2 (10^7cm^{-2})	$\omega_p^2 \cos Q\pi/2$ (10^7cm^{-2})	ω_f^2 (10^7cm^{-2})	β
TTF-TCNQ (300K)	13.0	8.4	9.1	0.85
TTF-TCNQ (60K)	15.9	9.5	12.3	0.56
TMTSF-DMTCNQ (300K)	12.5	8.8	8.5	1.08
HMTSF-TCNQ (300K)	20.2	8.0	13.0	0.59
DBTTF-TCNQCl ₂ (300K)	5.6	3.6	3.1	1.25
HMTSF-TNAP (300K)	12.8	7.8	10.3	0.50
TMTTF ₂ PF ₆ (300K)	7.9	5.6	5.8	0.91
TMTSF ₂ CIO ₄ (300K)	9.8	6.9	8.3	0.52
ET ₂ I ₃ (l) (300K)	8.3	5.9	5.2	1.29
ET ₂ I ₃ (l) (40K)	9.2	6.5	6.2	1.11

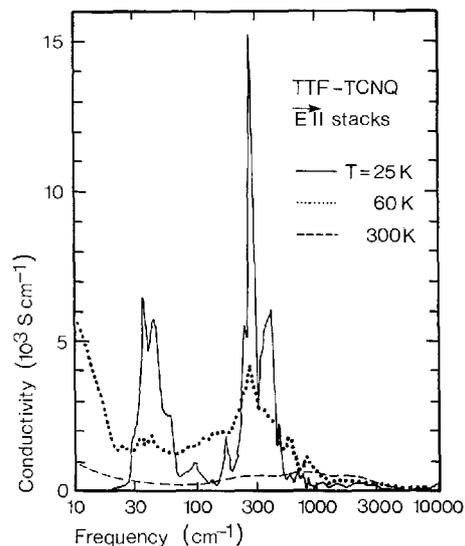
with independent information on the importance of correlation effects. The relatively big reduction of oscillator strength found indicates a considerable localization trend in the wavefunctions in full agreement with the enhancement in magnetic susceptibility generally observed (Torrance et al., 1977). The β -values, through Eq. (13), may prove useful in testing specific models for the wavefunctions in the intermediate correlation regime.

TTF-TCNQ: $2k_F$ - and $4k_F$ -instabilities

As the first good organic metal prepared, TTF-TCNQ remains one of the best characterized solids at all. Even so, many details remain controversial, one of them being the roles played by the two stacks. The physical and structural properties may briefly be summarized as follows:

The conductivity increases as T^{-2} or faster down to well below 100K (Cohen et al., 1974). At 60K $\sigma(T)$ is sharply peaked, and from 53K and down a cascade of phase transitions destroys the metallic character. Extensive structural studies (see, for example, Comes and Shirane, 1979) have shown considerable 1D diffuse scattering at $4k_F$ from 300K and down to the phase transitions. Below 150K appreciable $2k_F$ -scattering

Fig. 14. Frequency dependent conductivity of TTF-TCNQ at 25K, 60K and 300K. Notice the logarithmic frequency scale. (Tanner and Jacobsen, 1982).



develops. The character is 1D to 53K, where it becomes 3D. Several physical properties indicate that the 53K transition takes place on the TCNQ-stacks and that the TTF-stacks follow at a lower temperature (Schultz and Craven, 1979). It also appears that the TCNQ-stacks carry most of the conductivity in this material. The magnetic susceptibility is enhanced by a factor 2-3 over the Pauli value at room temperature and decomposition into individual stack contributions tends to show that the TTF-stacks have most of the susceptibility. While these results agree with the bandwidth estimates of Table IV they disagree with a recent NMR study (Takahashi et al., 1984).

Here we will discuss the infrared properties in general and in particular try to assign the $2k_F$ - and $4k_F$ -instabilities to individual stacks. The overall infrared properties have recently been studied by Jacobsen (1979), Tanner et al. (1981), and Tanner and Jacobsen (1982). The temperature dependence of $\sigma(\omega)$ is shown in Fig. 14 in the entire infrared range. Eldridge and Bates (1983) have subsequently studied the low temperature, far infrared spectrum by a different technique. While their results are in general agreement with those shown here, details in especially the sharpness of the features differ.

At 25K, $\sigma(\omega)$ displays a double peak structure with a low frequency band near 40 cm^{-1} and a very intense band near 300 cm^{-1} . The low frequency band contains about 5% of the total oscillator strength. In view of the observed $2k_F$ -superstructure and an energy gap estimated

from $\sigma_{dc}(T)$ of order 300 cm^{-1} , the intense band may be ascribed to single particle transitions across the gap in a $2k_F$ -Peierls semiconducting state, while the 30 cm^{-1} band is assigned to the Fröhlich (i.e. CDW) pinned mode.

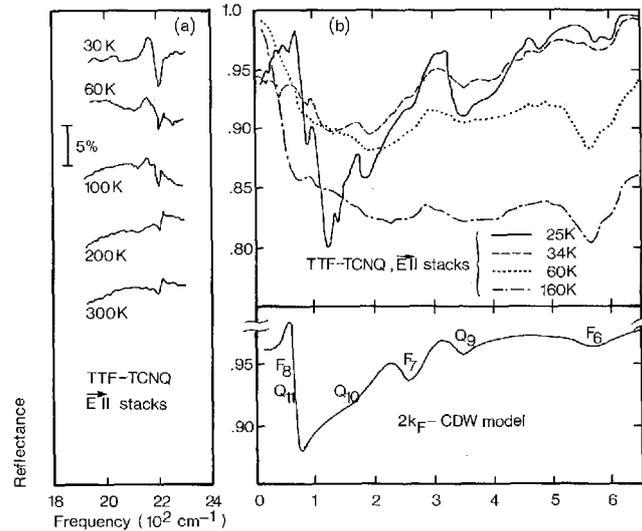
The CDW parameters estimated on basis of such an interpretation agrees with theoretical expectations (Tanner et al. 1981). However, a detailed fit of the spectrum in terms of the multiphonon theory previously employed is not yet possible.

The 60K spectrum shows a broadening of the 300 cm^{-1} structure and apparently the oscillator strength of the low frequency band has moved to zero frequency. Since $\sigma_{dc}(60K) \approx 10^4 \text{ S cm}^{-1}$, there is a sharp drop in $\sigma(\omega)$ in the millimetre and submillimetre range. Physically, this is consistent with a depinning of the CDWs at the phase transition. They appear to contribute to σ_{dc} and in the IR a pseudo-gap induced absorption, broadened by fluctuations, may be followed to temperatures above 100K. Going to 300K the features are much broadened and the maximum in $\sigma(\omega)$ has moved to about 800 cm^{-1} . The latter change can not be understood in terms of the $2k_F$ -instability.

Instead we may focus on the role of the $4k_F$ -instability which gives rise to the only detectable superstructure at 300K. It may be associated with only one or with both types of stack. As usual we expect the infrared properties of the $4k_F$ -CDW to consist of a maximum in $\sigma(\omega)$ corresponding to a pseudogap plus a number of sharper features near the A_g -vibrational modes. TCNQ has such a mode near 2200 cm^{-1} (C \equiv N stretch), which couples strongly to the electrons. Fig. 15(a) shows a high resolution study of the reflectance near this mode. At 300K and 200K the feature has the strength and shape of an ordinary IR active mode superimposed on the metallic response of the conduction electrons. From 100K and down to 30K the oscillator strength increases and the shape is inverted (the original structure is presumably hidden behind the new band). The low temperature spectrum is exactly what is expected from a Peierls distorted semiconductor. The point is that the temperature dependence follows that of the $2k_F$ -scattering. Thus the $4k_F$ -instability must (at least largely) take place on the TTF-stacks. This conclusion agrees with recent diffuse X-ray studies on irradiated TTF-TCNQ (Forró et al., 1984).

It is also corroborated by a careful inspection of the far IR reflectance, shown in Fig. 15(b). Being raw data, these curves should be reliable with respect to position and relative strength of fine structure. Also shown is a model calculation for a $2k_F$ -CDW system with $2\Delta = 300 \text{ cm}^{-1}$, using the

Fig. 15. Chain-axis reflectance of TTF-TCNQ. (a) Temperature dependence in the vicinity of the $C\equiv N$ stretch frequency. (b) Temperature dependence in the far infrared (Tanner et al., 1981). At bottom is shown a $2k_F$ -CDW model with identified A_g -mode features. Q refers to TCNQ, F to TTF.



known emv coupling constants of TTF and TCNQ (references from Table VII). The model mainly serves to identify the individual A_g -modes in the far infrared. It is again evident that TCNQ-modes (Q_{10} , Q_9) show a temperature dependence as that of the 2200 cm^{-1} mode, while the TTF-modes (F_6 , F_7) are different. F_6 corresponds to a strong, near temperature independent dip in the reflectance from 160K to 60K. Below 60K it moves slightly up in frequency and becomes weaker with a different shape. This behaviour may tentatively be assigned to a cross-over from a $4k_F$ - to a $2k_F$ -CDW state on the TTF-chains.

The infrared properties are clearly consistent with the idea that the $2k_F$ (ordinary) Peierls instability develops on the TCNQ-chains and perhaps at low temperature induces $2k_F$ -order on the TTF-chains as well. The $4k_F$ -instability in contrast develops on the TTF-chains. This also agrees with the bandwidth estimates of Table IV. The rather narrow TTF-band is susceptible to a correlation induced $4k_F$ -instability. The theory applied to the case of DBTTF-TCNQCl₂ is probably inappropriate, but qualitatively we may understand the 300K spectrum of TTF-TCNQ (see Fig. 11) as arising from a near Drude-like contribution from the TCNQ-stack and a contribution from the TTF-stack, which has a maximum near 800 cm^{-1} and emv interference effects in the $1400\text{--}1500\text{ cm}^{-1}$ range, where the important carbon skeleton modes are situated (compare Table VII). Interestingly, the rise in $\sigma(\omega)$ at low frequencies may arise from sliding $4k_F$ -CDWs, which here (in contrast to the case of DBTTF-

TCNQCl₂) experience no Coulomb scattering from oppositely charged density waves.

Conclusions

In the above study we have completed the following line of argumentation: It was shown empirically that the plasma edge Drude analysis of organic conductors yields reliable estimates for bandwidths. This in particular means that the plasmon frequency is largely insensitive to the often strong deviations from Drude behaviour encountered at low frequencies.

By comparing the bandwidth values with the integrated intraband oscillator strength, information on the importance of short range Coulomb correlations was extracted. It appears that all organic conductors have prominent »Coulomb holes«, but clear differences are found between good metals and moderate conductors. For the latter, models where double occupancy of sites is excluded have proven useful in explaining the semiconducting nature. In other materials, where Coulomb correlation effects appear equally strong, interchain interactions preserve the metallic character to low temperature, where even superconductivity may be found (as exemplified by ET₂I₃).

Finally, those molecular vibration modes, which couple to the electrons, were used as »fingerprints« to understand the nature of TTF-TCNQ, where apparently one chain-type shows much stronger correlation effects than the other.

ACKNOWLEDGMENTS. This work could not have been done without the continuing collaboration with K. Bechgaard, who supplied most of the samples studied.

Other samples were prepared by J. R. Andersen, I. Johannsen, M. Albjerg, L. Groth-Andersen, and H. Wang.

Some of the infrared measurements (far IR studies on TTF-TCNQ and all measurements on TMTSF₂PF₆) have been done jointly with D. B. Tanner, who also contributed to the interpretation. The author acknowledges helpful discussions with many collaborators and colleagues, among them especially K. Mortensen, H. J. Pedersen, K. Carneiro, R. Bozio, P. M. Grant, M. J. Rice, J. B. Torrance, J. M. Williams, and V. M. Yartsev. He also thanks G. Rindorf, H. Soling, and N. Thorup for expertise help in all matters regarding crystallography. M. H. Jørgensen and J. Schjær-Jacobsen were instrumental in the automation of the IR spectrometer, which saved much time subsequently.

This work has been partially financed by the Danish Natural Science Research Council and the NATO Research Grants Programme.

References

- Allen, P. B. (1971) *Phys.Rev. B* 3, 305.
- Andersen, J. R., Bechgaard, K., Jacobsen, C. S., Rindorf, G., Soling, H., and Thorup, N. (1978) *Acta Cryst. B* 34, 1901.
- Baeriswyl, D., Harbeke, G., Kiess, H., and Meyer, W. (1982) In *Electronic Properties of Polymers*, ed. by Mort, J. and Pfister, G. (Wiley, New York), p. 267.
- Bardeen, J., Cooper, L. N., and Schrieffer, J. R. (1957). *Phys.Rev.* 106, 162; 108, 1175.
- Bechgaard, K., Kistenmacher, T. J., Bloch, A. N., and Cowan, D. O. (1977). *Acta Cryst. B* 33, 417.
- Bechgaard, K., Jacobsen, C. S., and Andersen, N. H. (1978). *Solid State Commun.* 25, 875.
- Bechgaard, K., Jacobsen, C. S., Mortensen, K., Pedersen H. J., and Thorup, N. (1980). *Solid State Commun.* 33, 1119.
- Bechgaard, K., Carneiro, K., Olsen, M., Rasmussen, F. B., and Jacobsen, C. S. (1981a). *Phys.Rev. Lett.* 46, 852.
- Bechgaard, K., Carneiro, K., Rasmussen, F. B., Olsen, M., Rindorf, G., Jacobsen, C. S., Pedersen, H. J., and Scott, J. C. (1981b). *J. Am. Chem. Soc.* 103, 2440.
- Bernasconi, J., Rice, M. J., Schneider, W. R., and Strässler, S. (1975). *Phys.Rev. B* 12, 1090.
- Bloch, A. N., Cowan, D. O., Bechgaard, K., Pyle, R. E., Banks, R. H., Poehler, T. O. (1975). *Phys.Rev.Lett.* 34, 1561.
- Bosch, A., and van Bodegom, B. (1977). *Acta Cryst. B* 33, 3013.
- Bozio, R., and Pecile, C. (1980). In *The Physics and Chemistry of Low Dimensional Solids*, L. Alcácer, ed. (Reidel, Dordrecht), p. 165.
- Bright, A. A., Garito, A. F., and Heeger, A. J. (1974). *Phys.Rev. B* 10, 1328.
- Bryden, W. A., Kistenmacher, T. J., Poehler, T. O., Chappell, J. S., Emge, T. J., Lee, M. M., Lerstrup, K. A., Wiygul, F. M., Cowan, D. O., Stokes, J. P., and Bloch, A. N. (1984). *Phys.Rev. B*, to be published.
- Chaikin, P. M., Greene, R. L., Etemad, S., and Engler, E. (1976). *Phys.Rev. B* 13, 1627.
- Chesnut, D. B. (1966). *J.Chem.Phys.* 45, 4677.
- Cohen, M. J., Coleman, L. B., Garito, A. F., and Heeger, A. J. (1974). *Phys.Rev. B* 10, 1298.
- Comès, R., and Shirane, G. (1979). In *Highly Conducting One-Dimensional Solids*, J. T. Devreese, R. P. Evrard, and V. E. van Doren, eds. (Plenum Press, N.Y.), p. 17.
- Conwell, E. M. (1980). *Phys.Rev. B* 22, 1761.
- Delhaes, P., Coulon, C., Amiell, J., Flandrois, S., Toreilles, E., Fabre, J. M., and Giral, L. (1979). *Mol.Cryst. Liq.Cryst.* 50, 43.
- Duke, C. B., Lipari, N. O., and Pietronero, L. (1975). *Chem.Phys.Lett.* 30, 415.
- Eldridge, J. E., and Bates, F. E. (1983). *Phys.Rev. B* 28, 6972.
- Etemad, S., Penney, T., Engler, E. M., Scott, B. A., and Seiden, P. E. (1975). *Phys.Rev. Lett.* 34, 741.
- Ferraris, J., Cowan, D. O., Walatka, V., and Perlstein, J. H. (1973). *J. Am. Chem. Soc.* 95, 948.
- Forró, L., Bouffard, S., and Pouget, J. P. (1984). *J.Phys. (Paris) Lettr.* 45, L543.
- Fröhlich, H., (1954). *Proc. Royal Soc. (London)* A223, 296.
- Grant, P. M. (1983). *J.Phys. (Paris)-Colloq.* 44, C3-847.

- Greene, R. L., Mayerle, J. J., Schumaker, R., Castro, G., Chaikin, P. M., Etemad, S., and LaPlaca, S. J. (1976). *Solid State Commun.* 20, 943.
- Gutmann, F., and Lyons, L. E. (1967). *Organic Semiconductors* (Wiley, New York).
- Herman, F. (1977). *Physica Scripta* 16, 303.
- Holstein, T. (1954). *Phys.Rev.* 96, 535.
- Holstein, T. (1964). *Ann.Phys. (N.Y.)* 29, 410.
- Horovitz, B., Gutfreund, H., and Weger, M. (1975). *Phys.Rev. B* 12, 3174.
- Hubbard, J. (1978). *Phys. Rev. B* 17, 494.
- Huizinga, S., Kommandeur, J., Sawatzky, G. A. Kopinga, K., de Jonge, W. J. M. (1979). *Lecture Notes in Physics* 96, 45.
- Jacobsen, C. S. (1979). *Lecture Notes in Physics* 95, 223.
- Jacobsen, C. S., Mortensen, K., Andersen, J. R., and Bechgaard, K. (1978). *Phys.Rev. B* 18, 905.
- Jacobsen, C. S., Pedersen, H. J., Mortensen, K., and Bechgaard, K. (1980). *J.Phys.C.: Solid St.Phys.* 13, 3411.
- Jacobsen, C. S., Tanner, D. B. and Bechgaard, K. (1981). *Phys.Rev.Lett.* 46, 1142.
- Jacobsen, C. S., Tanner, D. B., and Bechgaard, K. (1983). *Phys.Rev. B* 28, 7019.
- Jacobsen, C. S., Johannsen I., and Bechgaard, K. (1984). *Phys.Rev.Lett.* 53, 194.
- Jacobsen, C. S., Williams, J. M., and Wang, H. H. (1985). *Solid State Commun.* 54, 937.
- Jerome, D., and Schulz, H. J. (1982). *Advances in Physics* 31, 299.
- Jerome, D., Mazaud, A., Ribault, M., and Bechgaard, K. (1980). *J.Phys. (Paris), Lett.* 41, L95.
- Kagoshima, S., Ishiguro, T., and Anzai, H. (1976). *J.Phys.Soc.Jpn.* 41, 2061.
- Kaminskii, V. F., Prokhorova, T. G., Shibaeva, R. P., and Yagubskii, E. B. (1984). *Pis'ma Zh.Eksp.Teor.Fiz.* 39, 15 (*JETP Lett.* 39, 17).
- Kistenmacher, T. J. (1978). *Ann.N.Y.Acad.Sci.* 313, 333.
- Kistenmacher, T. J., Phillips, T. E., and Cowan, D. O. (1974). *Acta Cryst. B* 30, 763.
- Koch, B., Geserich, H. P., Ruppel, W., Schweitzer, D., Dietz, K. H., and Keller, H. J. (1985). *Mol.Cryst.Liq.Cryst.*, 119, 343.
- Lerstrup, K., Lee, M., Wiygul, F. M., Kistenmacher, T. J., and Cowan, D. O. (1983). *J.Chem.Soc.Chem.Comm.* 294.
- Liautard, B., Peytavin, S. Brun, G., and Maurin, M. (1982). *J.Phys. (Paris)* 43, 1454.
- Lieb, E. H., and Wu, F. Y. (1968). *Phys.Rev.Lett.* 20, 1445.
- Lipari, N. O., Rice, M. J., Duke, C. B., Bozio, R., Girlando, A., and Pecile, C. (1977). *Int.J. Quantum Chem.: Quantum Chem.Symp.* 11, 583.
- Lyo, S. K. (1978). *Phys.Rev. B* 18, 1854.
- Lyo, S. K., and Gallinar, J.-P. (1977). *J.Phys.C: Solid St.Phys.* 10, 1696.
- Maldague, P. F. (1977). *Phys.Rev. B* 16, 2437.
- Mazumdar, S., and Bloch, A. N. (1983). *Phys.Rev.Lett.* 50, 207.
- Megtert, S., Pouget, J. P., Comès, R. Garito, A. F., Bechgaard, K., Fabre, J. M., and Giral, L. (1978). *J.Phys. (Paris)* 39, L-118.
- Mortensen, K. (1982). *Solid State Commun.* 44, 643.
- Mortensen, K., Jacobsen, C. S., Lindegaard-Andersen, A., and Bechgaard, K. (1983). *J.Phys. (Paris) Colloq.* 44, C3-1349.
- Mortensen, K., Jacobsen, C. S., Bechgaard, K., Carneiro, K., and Williams, J. M. (1985). *Mol.Cryst.Liq.Cryst.*, 119, 401.

- Mott, N. F. (1949). Proc.Phys.Soc. (London) A62, 416.
- Overhauser, A. W. (1960). Phys.Rev.Lett. 9,462.
- Peierls, R. E. (1955). *Quantum Theory of Solids* (Oxford Univ. Press), p. 108.
- Phillips, T. E., Kistenmacher, T. J., Bloch, A. N., and Cowan, D. O. (1976). J.Chem. Soc.Chem.Comm. 334.
- Pines, D. (1964). *Elementary Excitations in Solids*. (Benjamin, New York).
- Pouget, J. P. (1981). Chemica Scripta 17, 85.
- Pouget, J. P. (1984). Unpublished work.
- Rice, M. J. (1978). Solid State Commun. 25, 1083.
- Rice, M. J. (1979). Solid State Commun. 31, 93.
- Rice, M. J., and Strässler, S. (1973). Solid State Commun. 13, 125.
- Rice, M. J., Duke, C. B. and Lipari, N. O. (1975). Solid State Commun. 17, 1089.
- Rice, M. J., Pietronero, L., and Brüesch, P. (1977). Solid State Commun. 21, 757.
- Rice, M. J., Yartsev, V. M., and Jacobsen, C. S. (1980). Phys.Rev. B 21, 3437.
- Ritsko, J. J., Sandman, D. J., Epstein, A. J., Gibbons, P. C., Schnatterly, S. E., and Fields, J. (1975). Phys.Rev.Lett. 34, 1330.
- Schultz, A. J., Stucky, G. D., Blessing, R. H., and Coppens, P. (1976). J. Am. Chem. Soc. 98, 3194.
- Schultz, T. D., and Craven, R. A. (1979). In *Highly Conducting One-Dimensional Solids*, J. T. Devreese, R. P. Evrard, and V. E. van Doren, eds. (Plenum Press, N.Y.), p. 147.
- Slater, J. C. (1951). Phys.Rev. 82, 538.
- Sokoloff, J. B. (1970). Phys.Rev. B 2, 779.
- Soling, H., Rindorf, G., and Thorup, N. (1981a). Acta Cryst. B 37, 1716.
- Soling, H., Rindorf, G., and Thorup, N. (1981b). Unpublished work.
- Tajima, H., Yakushi, K., Kuroda, H., Saito, G., and Inokuchi, H. (1984). Solid State Commun. 49, 769.
- Takahashi, T., Jerome, D., Masin, F., Fabre, J. M., and Giral, L. (1984). J.Phys. C: Solid State Phys. 17, 3777.
- Tanaka, J., Tanaka, M., Tanaka C., Ohno, T., Takabe, T., and Anzai, H. (1978). Ann. N.Y. Acad.Sci. 313, 256.
- Tanner, D. B., and Jacobsen, C. S. (1982). Mol.Cryst.Liq.Cryst. 85, 137.
- Tanner, D. B., Jacobsen, C. S., Garito, A. F., and Heeger, A. J. (1976). Phys.Rev. B 13, 3381.
- Tanner, D. B., Cummings, K. D., and Jacobsen, C. S. (1981). Phys.Rev. Lett. 47, 597.
- Thorup, N., Rindorf, G., Soling, H., and Bechgaard, K. (1981). Acta Cryst. B 37, 1236.
- Torrance, J. B. (1979). Acc.Chem.Res. 12, 79.
- Torrance, J. B., Tomkiewicz, Y., and Silverman, B. D. (1977). Phys.Rev. B 15, 4738.
- Weyl, C., Engler, E. M., Bechgaard, K., Jehanno, G., and Etemad, S. (1976). Solid State Commun. 19, 925.
- Williams, P. F., and Bloch, A. N. (1974). Phys.Rev. B 10, 1097.
- Wooten, F. (1972). *Optical Properties of Solids* (Academic Press, New York).
- Yagubskii, E. B., Shchegolev, I. F., Laukhin, V. N., Kononovich, P. A., Karatsovnik, M. V., Zvarykina, A. V., and Buravov, L. I. (1984). Pis'ma Zh.Eksp.Teor.Fiz. 39, 12 (JETP Lett. 39, 12).
- Yamaji, K. (1985). Mol.Cryst.Liq.Cryst., 119, 105.

HANS UFFE PETERSEN
*Function of tRNA
in Initiation of Prokaryotic Translation*

ABSTRACT. This paper is a review of studies on structural and functional aspects of the interactions between the methionine accepting tRNAs and other macromolecules (enzymes, protein factors and ribosomes) in the cellular reactions preceding the formation of the first peptide bond during protein biosynthesis. Before discussing the recent research results, the problem will be introduced in a chronological order according to the discovery of the required molecular components and their functions. This introduction also aims at giving the reader an impression of the importance of the chosen subject in the field of molecular biology.

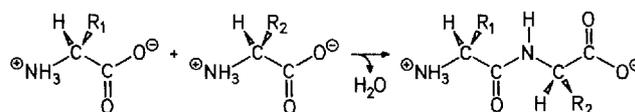
Division of Biostructural Chemistry,
Department of Chemistry, Aarhus University,
DK-8000 Aarhus C, Denmark

Introduction

All living cells contain a large number of different protein molecules. The exact number is not known, but estimates from the size of the genomes and two-dimensional polyacrylamide gel electrophoretic analysis of crude cell-lysates suggest the order of 2000 to 10000 different natural proteins.

These molecules consist of chains of amino acids linked together by peptide bonds as shown in Figure 1, where R_1 and R_2 are specific side-chains. All known proteins contain up to 20 different amino acids in the polypeptide chain – the primary sequence, which determines the three-dimensional structure and the function of each individual protein. The mechanism by which amino acids are linked together in the sequence specified by the nucleotide sequence of the gene is the biological translation. A four-letter “nucleotide alphabet” in DNA (and in messenger RNA) is translated into a 20-letter “amino acid alphabet” in protein molecules – a sequence of three nucleotides (codon) specifies each amino acid. The biological translation is a complicated process consisting of

Fig. 1. The formation of a peptide bond.



many steps and involving a number of specific macromolecular interactions. One of the most important components of this process is the tRNA molecule, which started its scientific career as a purely intellectual creation.

The suggestion and discovery of transfer RNA

After the elucidation of the DNA structure and the suggestion of the principles in the relation between DNA and protein in the early and mid-1950s, no one could envisage how nucleic acids could programme the synthesis of protein by direct structural interactions with amino acids. The first clue to this problem was the introduction by Francis Crick during the years 1955-57 of the "Adaptor Hypothesis" by which a small adaptor molecule would mediate between amino acids and a piece of nucleic acid which carried the genetic information for specifying the amino acid sequence during polymerisation into proteins. The Adaptor Hypothesis was almost simultaneously confirmed by the discovery of soluble RNA molecules carrying amino acids (Hoagland et al., 1959). The molecules were larger than expected by Crick, and turned out to be nucleic acids containing approximately 80 ribonucleotides and a molecular ratio of 25,000.

The first soluble RNA molecule (called sRNA – and since 1967 transfer RNA or tRNA) was sequenced by Holley et al. (1965) during the years 1958-65. The structure is shown in Figure 2A in the so-called cloverleaf representation. Many other tRNA molecules were soon sequenced, and a similar cloverleaf structure could be drawn for each molecule. The tRNA molecule can form intramolecular Watson-Crick base-pairs in four regions. One of these contains the terminals of the nucleotide chain, and the three others end in singlestranded loops. It is shown in Figure 2B where the amino acid attaches to the molecule and the "anti-codon" nucleotide sequence which "reads" the amino acid-specifying nucleotide sequence on messenger RNA is indicated at the bottom of the molecule.

The first three-dimensional structure of a tRNA molecule was obtained independently by the groups of Aaron Klug in Cambridge and Alexander Rich at M.I.T. in 1975. The folded structure (of phenyl-

alanine specific tRNA from yeast) is shown in Figure 3. Although minor and probably important differences may exist, the overall spatial structure is believed to be general in tRNA molecules.

The knowledge of the three-dimensional structure of the tRNA molecules is one of the prerequisites for the study of structural interactions between tRNA and proteins, which is one of the main subjects of this paper.

Protein biosynthesis starts at the amino terminal

At the time the tRNA molecules were discovered, only very limited knowledge existed about the molecular reactions in which it is involved during its functional cycle. The study of the components and mechanism of protein biosynthesis was just starting. However, it was becoming clear that the process took place on sub-cellular particles – first called microsomes and in 1958 named ribosomes by Richard Roberts.

Then it was a crucial question in which direction the reaction of amino acid polymerisation proceeded. If we look at Figure 1, would R_1 or R_2 be

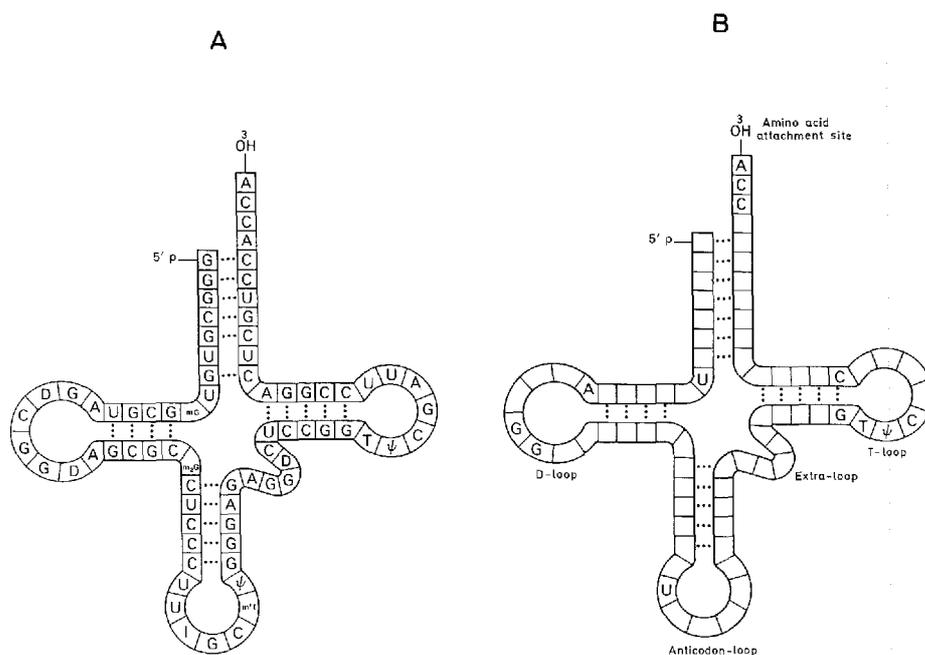
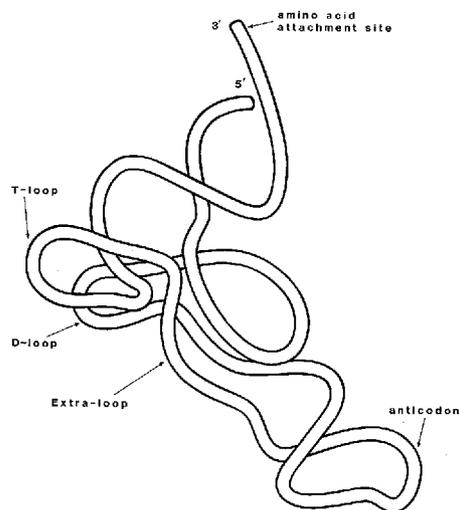


Fig. 2A. Cloverleaf representation of the nucleotide sequence of tRNA^{Ala} from yeast.
B: Common features of tRNA molecules.

Fig. 3. Schematic diagram of the three-dimensional structure of tRNA^{Phe} from yeast. Based on the X-ray crystallographic data of Ladner et al., 1975.



the first amino acid in the polypeptide – in other words would the initiating amino acid become the N- or C-terminal of the molecule?

The first answer to this question came from a study of the biosynthesis of haemoglobin by Bishop *et al.* (1960), who found that the polypeptide chain growth began with the N-terminal amino acid. This means that R₁ in Figure 1 is the terminal amino acid in a protein starting by this reaction. Shortly after, Goldstein and Brown (1961) showed that *E. coli* proteins are synthesized beginning with the N-terminal amino acid.

N-terminal amino acid = methionine

By that time, other workers were studying sequences of amino acids in proteins. From the work of Moriwitz and Spaulding (1958), it seemed that one particular amino acid appeared more than statistically in the N-terminal position of *E. coli* – namely methionine.

That methionine is found much more than randomly as the N-terminal amino acid of *E. coli* proteins was confirmed by Waller and Harris (1961) – for ribosomal proteins – and finally by Waller (1963) – for total *E. coli* proteins. This led to believe that proteins are synthesized by a mechanism which specifically incorporates methionine as the first and N-terminal amino acid. Twenty years ago, the picture of the molecular mechanism of protein biosynthesis can be presented as shown in Fig. 4 (Watson, 1964).

The genetic information in the chromosomal DNA (in the form of a fourletter nucleotide sequence) is transcribed into messenger RNA by

the enzyme RNA polymerase. The mRNA is then bound to a ribosome and by adaptor molecules – aminoacyl-tRNAs – translated into a 20-letter amino acid sequence in a polypeptide chain.

It was known that special protein factors (transfer enzymes) were needed to stimulate the overall process of translation and that energy was supplied by GTP, but the knowledge specifically concerning the mechanism by which the process started was clearly expressed by Watson: “Almost no hints exist about chain initiation”.

The first important discovery of initiation specific molecules came from the Cambridge laboratory of Perutz.

Initiator tRNA

In their work with sulfur containing amino acids (easily labelled with high specific radioactivity) and their interaction with tRNAs from *E. coli*, Marcker and Sanger (1964) found that [³⁵S]-methionyl-tRNA upon mild hydrolysis (which splits off the amino acid as shown in Fig. 16) gave two radioactive spots on electrophoretic analysis – one from methionine and another which was identified as N-formyl-methionine. Shortly after Marcker (1965) showed that two different classes of methionyl-tRNA exist of which one could be N-formylated in the methionine by an *in vitro* system, whereas the other could not. The two species were named Met-tRNA₁ (not formylatable, now called tRNA_m^{Met}) and Met-tRNA₂ (formylatable, now called tRNA_f^{Met}) and the role of formyl-methionyl-tRNA as initiator tRNA in protein synthesis was the authors' immediate suggestion. The cloverleaf structures of the two tRNAs are

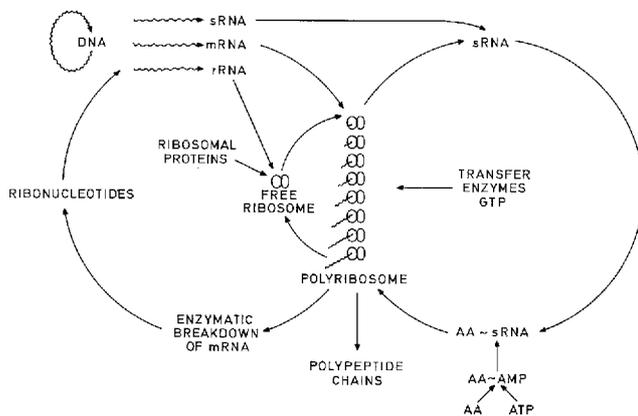


Fig. 4. Schematic view of protein synthesis as of March 1964 (Watson, 1964).

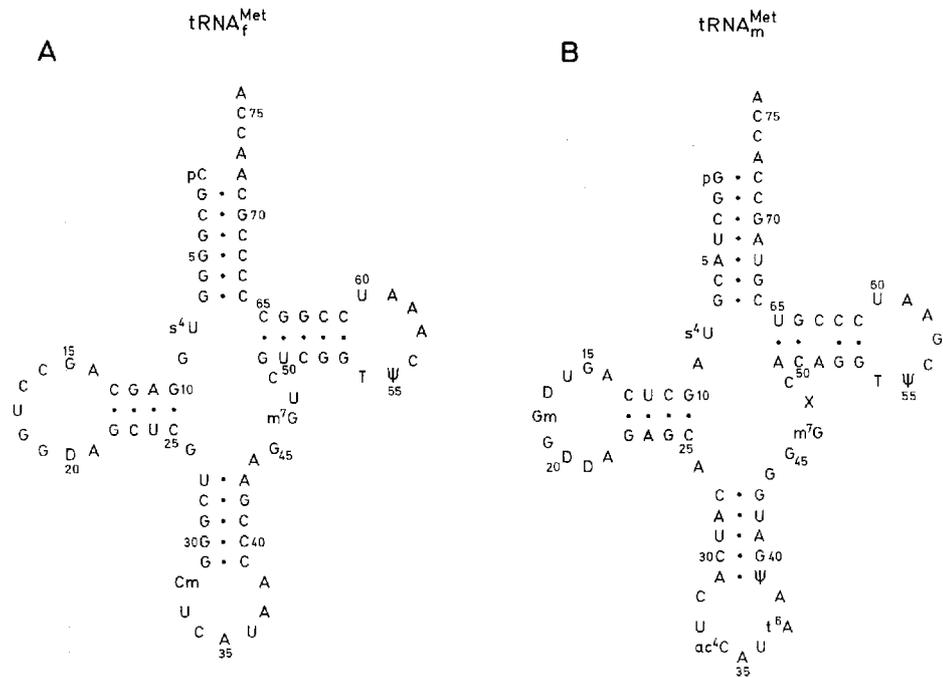


Fig. 5A. Cloverleaf representation of the nucleotide sequence of initiator $tRNA_f^{Met}$ (Dube et al., 1968).

B. Cloverleaf representation of the nucleotide sequence of elongator $tRNA_m^{Met}$ (Cory et al., 1968).

shown in Fig. 5. Similarities and differences in the nucleotide sequences will be discussed further in subsequent paragraphs. In an *in vitro* system similar to the one developed by Nirenberg and Matthaei (1961), Clark and Marcker (1966a) showed that N-formyl-methionyl- $tRNA_f^{Met}$ exclusively incorporated formyl-methionine into the N-terminal of a growing polypeptide chain (in response to poly(U,G)), similarly the same authors showed (1966b) that the non-formylatable $Met-tRNA_m^{Met}$ incorporated methionine into internal positions of polypeptides and not into the N-terminal. This was considered as a proof that $tRNA_f^{Met}$ is the true initiator tRNA in *E. coli*.

Aminoacylation of $tRNA_f^{Met}$

The tRNA molecules carry the amino acids from the cell cytoplasm to the ribosome – the site of protein biosynthesis. Now the question arises: How are the correct amino acids linked to the tRNA molecules? This is

an enzymatic process, which requires catalysis by an aminoacyl-tRNA synthetase specific for each amino acid.

The methionyl-tRNA synthetase aminoacylates both tRNA_m^{Met} and tRNA_f^{Met}. The enzyme has been extensively characterized by the group of Sylvain Blanquet at the Ecole Polytechnique in Paris (Blanquet *et al.*, 1973, 1976). The native protein contains two identical polypeptide chains (type α_2) each with an M_r of 76,000.

The monomer can easily be split by trypsin, resulting in a 64 k dalton fragment, which retains the activity of aminoacylation (Dessen *et al.*, 1982). This fragment has been crystallized, and the X-ray crystallography group of Risler in Gif-sur-Yvette has determined the crystal structure at 2.5 Å resolution (Zelwer *et al.*, 1982).

The aminoacylation reaction catalyzed by Met-tRNA synthetase is shown in Fig. 8 (Reactions 1 and 2).

Formylation of initiator tRNA

Marcker (1965) showed that the formylation in the α -amino group of methionine takes place after the aminoacylation of tRNA_f^{Met} and found that N¹⁰-formyl-tetra-hydrofolate was a good formyl-donor in the crude *in vitro* system. The formylating enzyme was purified by Dickerman *et al.* (1967), who confirmed that N¹⁰-fTHF was the formyl-donor. Therefore, the enzyme was named: N¹⁰-formyl-tetrahydrofolate:L-methionyl-tRNA N-formyltransferase (E.C.2.1.2.9). This means that the formylation of Met-tRNA_f^{Met} is coupled to the C-1 metabolism as shown in Fig. 6.

The transformylase has been extensively purified by Kahn *et al.* (1980) who showed that at physiological ionic conditions, the enzyme has a significantly higher affinity to charged initiator tRNA_f^{Met} as compared to uncharged tRNA_f^{Met} and to other tRNAs.

From the reactions of aminoacylation and formylation (and as it will be discussed later, also from the interaction with the elongation factor Tu), it seems clear that the tertiary structure of the initiator tRNA must be partially identical to that of tRNA_m^{Met} but also partially different from all other tRNAs, and this specific tertiary structure seems sensitive to aminoacylation.

Not only in *E. coli* but also in cell organelles of eukaryotic organisms (chloroplasts and mitochondria) is the protein synthesis initiated by fMet-tRNA (Galper and Darnell, 1969). This led Marcker (1969) to propose the general rule that prokaryotic cells and eukaryotic cell organelles use N-formyl-methionyl-tRNA as polypeptide chain initiator

Formylation of prokaryotic Met-tRNA_f^{Met} is not indispensable.

In vivo evidence against the general rule of formylation

However, formylation of initiator tRNA methionine in prokaryotic organisms is not general – that cells may grow without formylation has been shown in different organisms in four principal classes of cases:

- 1: organisms which cannot synthesize the formyl-donor *de novo* may grow in the absence of formylation
- 2: organisms which contain no transformylase enzyme grow without formylation
- 3: *E. coli* can grow when formylation is inhibited
- 4: mutants of *E. coli* have been isolated which grow normally in the presence of an inhibitor of formylation.

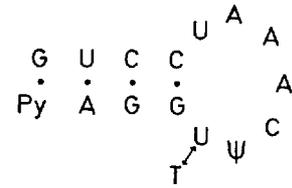
An example of each of these cases will be described briefly:

- 1: The organism *Streptococcus faecalis* R is (in contrast to *E. coli* and other prokaryotes) not able to synthesize folic acid *de novo*, but it was found that the cells could grow in the absence of folic acid when the substrate was supplemented with the folic acid metabolites serine, methionine, thymine and a purine base – see Fig. 6. At such conditions, the formyl-donor N¹⁰-formyl-tetrahydrofolate cannot be synthesized, and formylation of the initiator tRNA must (if there is a formylation) follow other metabolic pathways (Samuel *et al.*, 1970). Indeed, no formylation or other blocking of the α -amino group of Met-tRNA methionine took place in the folate-free medium, although in the presence of folic acid, most of the Met-tRNA_f^{Met} was formylated *in vivo* (Samuel *et al.*, 1972). Later, the authors showed that the tRNA_f^{Met} from folate-free cells was altered as compared to the initiator tRNA isolated from cells grown in the presence of folic acid – the ribothymidylic acid normally found in the T ψ C loop was not found in the “folate-free” tRNA, see Figs 5A and 7. So it seemed that a structural change in the initiator tRNA could compensate for the lack of formylation (Samuel & Rabinowitz, 1974).

However, it was found that the methylase responsible for the formation of rT in the tRNA in *S. faecalis* requires a folate coenzyme as methyl donor. This means that the absence of formylation and of rT in the tRNA of “minus folate” cells are both (and possibly independently) induced by the folate deprivation (Delk & Rabinowitz, 1975).

Rabinowitz and coworkers did not study the *in vitro* binding of fMet-tRNA_f^{Met} from plus and minus folate *S. faecalis* in crossed experi-

Fig. 7. The T-arm of the initiator $tRNA_f^{Met}$ from *S. faecalis*. The ribothymidylic acid present in +folate cells is replaced by uridine in -folate cells (Samuel and Rabinowitz, 1974).



ments with ribosomes (and initiation factors) isolated from the two growth conditions. It can therefore not be excluded that the ribosomes (and/or the initiation factors) play an important role for the ribosomal binding of the unformylated (and undermethylated) initiator tRNA from “minus folate” *S. faecalis* stimulated by initiation factors (and thus in the ability of *S. faecalis* cells to grow in the absence of formylation).

- 2: The strongly halophile *Halobacterium cutirubrum* was studied by White & Bayley (1972) who found that this organism does not synthesize the transformylase enzyme needed for the formylation of $Met-tRNA_f^{Met}$, although the methionine accepting tRNA isolated from *H. cutirubrum* could be separated into two species of which one was formylatable by *E. coli* transformylase.

As the organism requires extreme high salt concentrations, it is likely that conformational changes of the tRNA (and/or the ribosomes) may occur as compared to a hypothetical low-salt structure. Such salt induced structural changes may explain the ability of *H. cutirubrum* to grow without formylation of the initiator tRNA.

- 3: The antibiotic trimethoprim is an inhibitor of dihydrofolate reductase, the enzyme which reduces dihydrofolate (DHF) to tetrahydrofolate (THF) – see Fig. 6. Harvey (1973) has studied the effect of trimethoprim on the growth of *E. coli*. He found that *E. coli* wild type (B/r) cells could grow exponentially at a reduced rate in the presence of trimethoprim when the folic acid metabolites were added. Under these conditions, no formylation of $Met-tRNA_f^{Met}$ took place.

Harvey found that at the growth conditions where the initiator tRNA was unformylated, a two-fold increase in the ratio ribosomal subunits:70S ribosomes arised. As will be discussed further in the following sections, the increased level of ribosomal subunits may be involved in the explanation of the ability of the cells to grow without formylation of $Met-tRNA_f^{Met}$.

- 4: The isolation of *E. coli* mutants able to grow normally without for-

mylation (Danchin, 1973) is an additional support for the idea that formylation of initiator Met-tRNA is not general and vital for prokaryotic organisms.

Danchin isolated *E. coli* mutants on media containing the C-1 metabolites supplemented with trimethoprim, and found quite surprisingly that among these mutants were strains resistant to streptomycin. Such rpsL mutants are known to be point mutations in the ribosomal protein S12. Therefore, it seems clear that the ability to grow in the absence of formylation in some way is linked to alterations in the structure of the ribosomes.

Another *E. coli* mutant was isolated by Baumstark *et al.* (1977) which was able to grow without formylation of the initiator tRNA. By the same method as previously used by Danchin (1973), a K12 mutant was selected in the presence of sulfathiazole and trimethoprim. The authors showed that the initiator tRNA_f^{Met} of this mutant contained a reduced (but finite) amount of ribothymidine in the T Ψ C loop of the tRNA and concluded that this alteration in the initiator tRNA structure was possibly the reason why the mutant could grow without formylation.

However, their *in vitro* f2 directed protein synthesis did not include crossed experiments with mutant ribosomes and unformylated Met-tRNA_f^{Met} isolated from the parental strain cells. This would be required to exclude that the ability to grow in the absence of formylation arises from altered ribosomes. In fact, the mutant of Baumstark is not a single mutation but derived from a parental strain which was streptomycin resistant (altered in the ribosomal protein S12).

E. coli mutants containing unmethylated uridine instead of rT in the tRNA have been isolated by Björk & Isaksson (1970). In order to test the hypothesis of Baumstark and coworkers, trimethoprim was added to the growth medium of the mutants (and the C-1 metabolites). Under such conditions no formylation of Met-tRNA_f^{Met} could take place, and it would show if the lack of rT was sufficient to account for growth in the absence of formylation. The cells were not able to grow under such conditions (Danchin and Isaksson, personal communication). Thus it seems clear that the other mutations in the strain isolated and studied by Baumstark, in addition to the mutation in the methylase, must play a role in the ability of these cells to grow without formylation of the initiator Met-tRNA.

In vitro evidence against the general rule of formylation

Looking back at the first experiments *in vitro* with the initiator tRNA there were indications that formylation might not be indispensable. Clark & Marcker (1966a) did not find any significant difference in the activity of Met-tRNA_f^{Met} whether formylated or not, and just before the discovery of the initiation factors in 1966, Bretscher & Marcker (1966) had shown that at 10-20 mM magnesium, both formylated and unformylated Met-tRNA_f^{Met} could bind to ribosomes and react with puromycin. They concluded that the only role of the formyl group was to increase the rate of initiation.

This led Clark and coworkers to believe that the specificity as initiator tRNA lay in the tRNA structure and not in the presence of the formyl group on the methionine. Therefore, the two methionine accepting tRNAs were sequenced in the hope that the nucleotide primary structure could reveal some characteristic differences between the two tRNAs (on the one hand and between tRNA_f^{Met} and all other tRNAs on the other).

In the two cloverleaf structures shown in Fig. 5, one observes a number of differences in nucleotides at the individual positions. However, an intensive comparison of the tRNA_f^{Met} sequence with other tRNAs shows that all individual nucleotides and local sequences of tRNA_f^{Met} can be found in other tRNAs. Only one difference in the primary structure seems to be characteristic for tRNA_f^{Met} – namely the fact that the 5'-terminal nucleotide (cytidine) is not involved in a Watson-Crick base-pair which is the case for the 5'-terminal nucleotide of all other tRNAs.

Although this may be very important, it is unlikely that it accounts for all the specificities which are found for the initiator tRNA as compared to tRNA_m^{Met} and all other tRNAs, especially as this difference is not found in eukaryotic cells which also use a formylatable (by *E. coli* transformylase) Met-tRNA species as initiator tRNA.

Other workers then found that formylation was required for the translation *in vitro* of natural mRNA (from phage f2) at 5 mM magnesium (Eisenstadt & Lengyel, 1966), and after the discovery of the initiation factors and the establishment of their optimal salt conditions (5-10 mM magnesium), many *in vitro* experiments suggested that fMet-tRNA_f^{Met} was the correct initiator tRNA as only the formylated tRNA species could bind to 70S ribosomes strongly stimulated by initiation factors (the concept of "enzymatic binding") (Grunberg-Manago *et al.*, 1969; Drews *et al.*, 1973).

In most of these experiments, 70S ribosomes were used for studying

the initiation factor stimulated binding. As the initiation mechanism was supposed to involve the formation of a 30S initiation complex with mRNA and initiator tRNA (see Fig. 8), one could ask whether 30S subunits would behave differently from the 70S ribosomes in such *in vitro* binding experiments. A first indication of such a difference can be seen from the work of Grunberg-Manago *et al.* (1969). In one experiment, 30S subunits were used, and indeed, a small stimulation (in absolute scale, but 50% relative) by the initiation factors of the binding of unformylated Met-tRNA_f^{Met} at 5mM magnesium is observed.

In the work of Samuel & Rabinowitz (1974) previously described, the initiator tRNA from "minus-folate" *S. faecalis* cells was extracted and purified. In the *in vitro* ribosomal binding assay, the authors showed that initiation factors stimulated the unformylated and the formylated form of this Met-tRNA_f^{Met} to the same degree. (This IF-stimulated binding was seen at Mg⁺⁺ concentrations up to 30 mM.)

Together these results indicate that although the formylation may increase the rate of interaction of the initiator tRNA with ribosomes, the enzymatic binding of Met-tRNA_f^{Met} to ribosomes or ribosomal subunits does not absolutely require formylation.

More recently, Rich and collaborators have used the single-strand specific endonuclease S1 to study structural characteristics of initiator tRNAs compared to elongator tRNAs (Wrede *et al.*, 1979). They found that S1 cleaved at two distinct sites in the anticodon loop of three different initiator tRNAs (from *E. coli*, yeast and mammalian), whereas elongator tRNAs were cut at four sites. The authors suggest that the specificity is caused by the three G-C base-pairs of the anticodon arm (see Fig. 5), which are common in the initiator tRNAs and different from the elongator tRNAs included in the experiments by Rich & Wrede. However the initiator tRNAs in Bean chloroplasts and in yeast mitochondria contain only two G-C base pairs in these positions. The nucleotide at position 29 is here A in stead of the G in *E. coli* tRNA_f^{Met} (Fig. 5A). Furthermore, at least two examples are known of elongator tRNAs which contain three G-C base pairs at position 29-41, 30-40 and 31-39, namely tRNA^{Met} from Bovine liver mitochondria and tRNA^{Ser} from *Halobacterium vol.* (Sprinzl & Gauss, 1984). Therefore it seems clear that no single local nucleotide sequence in the anticodon stem of the initiator tRNA_f^{Met} is determining its specific function. Indeed as it will be discussed in later paragraphs, there are indications that unformylated Met-tRNA_f^{Met} can function as an elongator tRNA.

As it seems unlikely that formylation should cause a drastic change in

the structure of the anticodon stem, the results of Rich *et al.* are in agreement with our results presented in the following sections, that formylation is not the determining factor but that the tertiary structure of the initiator tRNA most probably is one of the most important features in the functional specificity of fMet-tRNA_f^{Met}. We have studied the accessibility of different regions of tRNA_f^{Met} and tRNA_m^{Met} to a number of ribonucleases and compared the results in the presence and absence of proteins. These results will be discussed in detail in subsequent paragraphs. Other important elements which will be discussed in the following is the presence *in vivo* of translation factors. The fact that unformylated Met-tRNA_f^{Met} binds to the elongation factor EF-Tu (see later) indicates that at extreme growth conditions, some of the initiator tRNAs might in fact function in the elongation step of translation.

Buckingham and collaborators have found that initiation factors stimulate the binding of formylated or acetylated Trp-tRNA^{Trp} to 30S or 70S ribosomes in the presence of poly(U,G). A number of other tRNAs have been tested for similar initiator activity with negative results (Leon *et al.*, 1979). These results indicate some structural similarity between tRNA_f^{Met} and tRNA^{Trp}, although such similarity does not appear from a comparison of the nucleotide sequences of the two tRNAs. One point which may seem rather odd is that these two tRNAs, alone among the 21 different amino acid acceptors, do not show degeneracy in their corresponding codons, namely AUG and UGG respectively, are codons for tRNA_f^{Met} and tRNA^{Trp} (with the exception that the valine codon GUG and the normal codon for isoleucine AUU in a few cases have been found as initiator codons).

Finally, the amino acid part of fMet-tRNA_f^{Met} seems to be of minor importance for the functional specificity as initiator tRNA. Giegé *et al.* (1973) showed that the methionine of fMet-tRNA_f^{Met} could be replaced by valine or phenylalanine without any change in the initiation factor-stimulated binding to ribosomes. The puromycin reactivity too was unchanged by the mischarged f-aa-tRNA_f^{Met}s.

As a conclusion, it seems that the ribosome-mRNA interaction with the initiator tRNA must be considered when determining the specificity of a tRNA as initiator. The proper selection of the initiator tRNA may at least in a final step be controlled by the initiator codon of the mRNA, and although formyl-Trp-tRNA^{Trp} can bind to the P-site of the 30S or 70S ribosome (*in vitro*), the use as initiator tRNA *in vivo* is prevented by the lack of formylation and by the mis-matching of the tRNA^{Trp} anticodon and the initiator codon of the mRNA.

So the total recognition between the components in the initiation complex is most probably a combination of the mRNA-aa-tRNA, the mRNA-30S, the aa-tRNA-IF2 and the 30S-aa-tRNA interactions, and although one or more of these interactions might proceed by species other than (f)Met-tRNA_f^{Met}, the final complex will not be stable in such cases.

A summary of the reactions involving the initiator tRNA_f^{Met} during the initiation of translation

A short summary of the present view of the mechanism of polypeptide chain initiation follows. Fig. 8 is a schematic representation of the steps involved in the initiation process (Petersen *et al.*, 1984b).

After the aminoacylation (reaction 1+2) and formylation (reaction 3) of tRNA_f^{Met}, a pre-initiation complex between fMet-tRNA_f^{Met} and initiation factor IF2 is formed (reaction 4). GTP may be involved in the formation of a ternary complex fMet-tRNA_f^{Met}:IF2:GTP, but no direct evidence for such a complex has yet appeared; a later section discusses this subject in more detail.

The next step involves the ribosome and in particular the 30S ribosomal subunit. The 70S particle exists in equilibrium with the 30S and 50S subunits (reaction 11). *In vitro* experiments have shown that temperature, pH and especially ionic conditions are parameters of importance for this equilibrium. Thus increased concentration of potassium or decreased concentration of magnesium promotes dissociation (see also Fig. 10). However, in the cell, the equilibrium is supposed to be controlled by special proteins, the initiation factors IF1 and IF3. IF1 increases the rates of dissociation and association, and IF3 prevents reassociation by binding to the 30S subunit. IF3 is thus acting as an anti-association factor, and in the presence of IF3, IF1 acts as a dissociation factor, thus in collaboration, the two factors shift the equilibrium towards the 30S + 50S state (Grunberg-Manago, 1980).

The sequence in the following steps, the binding of mRNA and fMet-tRNA_f^{Met} to the 30S subunit, is not known. There exist indications for both molecules being bound prior to the other. Therefore, in Fig. 8, the binding of both components is shown simultaneously.

In the review previously mentioned, Watson (1964) introduced the concepts of two distinct ribosomal sites for interaction with mRNA and aminoacyl-tRNA – now called the P- and A-sites. The initiator region

Table 1: Functional differences between tRNA_f^{Met} and other tRNAs

	Codon specificity	Interaction with proteins and ribosomes					
		Met-tRNA synthetase (E.C.6.1.1.10)	Transformylase* (E.C.2.1.2.9)	Initiation factor IF2	Elongation factor EF-Tu	70S ribosomal A-site	P-site
tRNA _f ^{Met}	AUG (GUG) [#] , (AUU) [#] , (UUG) [#]	⊕	+	⊕	⊕	±	⊕
tRNA _m ^{Met}	AUG	⊕	-	-	⊕	+	-
other tRNAs	not AUG	-	-	-	+	+	-

* after charging with the cognate amino acid

[#] shown to stimulate Met-tRNA_f^{Met} binding to ribosomes *in vitro* (Clark & Marcker, 1966b)

found as initiator codon in mRNA (Steitz, 1980; Sacerdot *et al.*, 1982).

⊕ complexes discussed in this paper.

only recently been investigated. Such studies are partly the subject of the following sections. The functional differences between tRNA_f^{Met} and other tRNAs are shown in Table 1. The proteins and ribosomal interactions involving the different tRNAs are indicated with encircled symbols at complexes which are discussed in the subsequent sections.

In vitro studies of the 30S and 70S ribosomal Met-tRNA complex

The initiation of translation is believed to proceed via the formation of a complex between mRNA, 30S ribosomal subunit and formylmethionyl-tRNA_f^{Met} as shown in Fig. 8. Because of the several exceptions to the general rule that the initiator tRNA methionine must be formylated, we wanted to study the *in vitro* formation of the 30S initiation complex. We have measured the binding of initiator Met-tRNA_f^{Met}, formylated and unformylated, to 30S ribosomal subunits, and we have studied the effects of different molecular components on this binding (Petersen *et al.*, 1976a).

Table 2 shows that crude initiation factor extracts stimulate the binding of both tRNA species in the presence of the synthetic messenger poly(A,G,U). In addition, it was shown that purified initiation factors

Table 2: Effect of different messengers on the binding of fMet-tRNA_f^{Met}, Met-tRNA_f^{Met}, and Met-tRNA_m^{Met} to 30S ribosomal subunits in the absence and in the presence of initiation factors and GTP

Messenger	GTP	Crude IF	[Mg ⁺⁺], mM	Relative molar binding to 30S ribosomal subunits		
				fMet-tRNA _f ^{Met}	Met-tRNA _f ^{Met}	Met-tRNA _m ^{Met}
No	+	-	5	1	3	2
	+	+	5	4	2	1
ApUpG	+	-	5	3	2	2
	+	+	5	27	7	3
R17RNA	+	-	5	1	6	1
	+	+	5	74	13	1
Poly(A,G,U)	+	-	5	21	17	3
	+	+	5	183	35	8
	-	-	5	12	18	
	-	+	5	112	68	
	-	-	15	46	45	
	-	+	15	127	107	
	-	-	35	49	52	
	-	+	35	117	95	

(Petersen *et al.* 1976a).

had the same effect. This stimulation also exists in the presence of the trinucleotide codon for methionine ApUpG and in the presence of the natural messenger RNA from phage R17. No significant binding of the elongator Met-tRNA_m^{Met} is found under any of these conditions.

Subsequently, preassociated 70S (tight couples) ribosomes were used in similar binding experiments, and as seen in Table 3, unformylated initiator Met-tRNA_f^{Met} binds to 70S ribosomes in the absence of initiation factors. At low magnesium concentration, this binding is almost completely inhibited when initiation factors are present during the incubation.

Contrastingly, formyl-Met-tRNA_f^{Met} does not bind significantly to 70S ribosomes at low magnesium concentration. The addition of initiation factors shows the well-known stimulation of the binding of fMet-tRNA_f^{Met}.

Furthermore, Tables 2 and 3 show that GTP inhibits the binding of the unformylated initiator tRNA to both 30S and 70S ribosomes, both in the

Table 3: Effect of GTP on the binding of Met-tRNA_f^{Met} and fMet-tRNA_f^{Met} to 70S ribosomes in the absence and in the presence of initiation factors

Addition	Crude IF	pmol	
		Met-tRNA _f ^{Met}	fMet-tRNA _f ^{Met}
none	–	1.89	0.15
GTP	–	1.64	0.50
none	+	0.07	0.24
GTP	+	0.04	1.85

Incubation mixtures contained: 3.8 pmol of Met-tRNA_f^{Met} or 2.8 pmol of fMet-tRNA_f^{Met}, 15 pmol of 70S ribosomes and, where indicated, 1 mM GTP and 49 µg of crude IF. Magnesium acetate, Tris-HCl buffer, ammonium chloride, and poly(A,G,U) were in the amounts described in the legend to figure 9 (Petersen *et al.*, 1976b).

absence and in the presence of initiation factors, whereas this highenergy nucleoside triphosphate, as is well-established, stimulates the binding of fMet-tRNA_f^{Met} in all cases. This may indicate that GTP binds directly to the 30S subunit and not through a complex with IF2.

To test whether the Met-tRNA_f^{Met} bound to 30S ribosomal subunits is bound at the ribosomal P-site, 50S subunits were added after the binding to 30S, and subsequently puromycin was added. This antibiotic resembles the 3'-end of aminoacyl-tRNA and reacts with aminoacyl or peptidyl groups of P-site bound aa-tRNA or peptidyl-tRNA.

The results are seen in Figure 9 which shows that at least a part of the Met-tRNA_f^{Met} which is bound to 30S ribosomal subunits is bound at a site which after addition of the 50S ribosomal subunit corresponds to the ribosomal P-site. About 50% of the totally bound Met-tRNA_f^{Met} transfer methionine to puromycin.

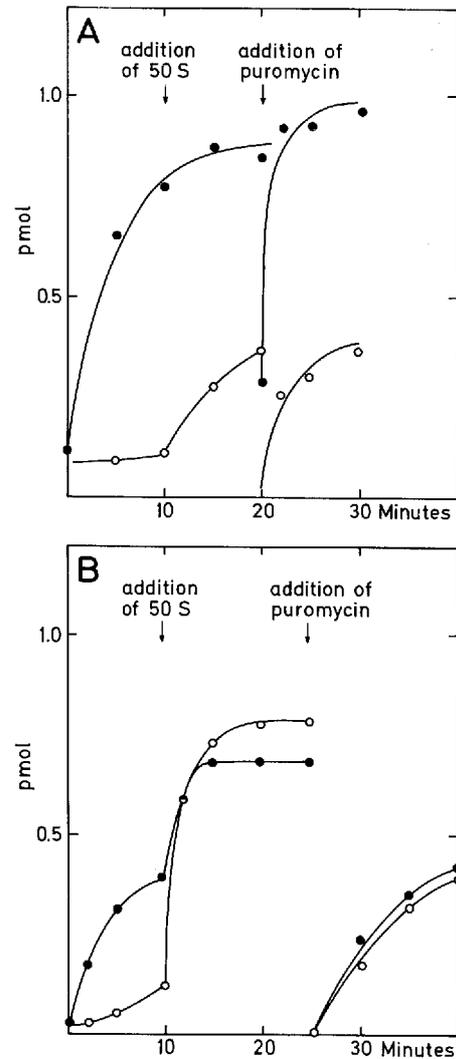
The results show that formylation does not seem to be needed for the formation of a 30S initiation complex. Unformylated Met-tRNA_f^{Met} is recognized by initiation factors in the presence of 30S ribosomal subunits and bound like fMet-tRNA_f^{Met} at a puromycin reactive site (measured after addition of 50S ribosomal subunits).

As it will be discussed in a later section, this recognition may be caused by ribosome-bound initiation factors as free IF2 does not seem to have any significant affinity to Met-tRNA_f^{Met}.

This indicates that it is the structure of the initiator tRNA_f^{Met} – and not the formyl group – which is recognized by the 30S ribosomal subunit in the presence of initiation factors.

Fig. 9A. Binding of $fMet-tRNA_f^{Met}$ to 30S ribosomal subunits, effect of 50S subunits, and puromycin sensitivity in the absence (\circ) and presence (\bullet) of initiation factors. Incubation mixtures ($50\ \mu\text{l}$) contained, at time 0, 15 pmol of 30S subunits, 50 mM Tris-HCl (pH 7.4), 0.11 A_{260} unit of poly(A,G,U), 5 mM magnesium acetate, 50 mM ammonium chloride, 1 mM GTP and 2.1 pmol $fMet-tRNA_f^{Met}$, and where indicated, 26 μg of crude initiation factor extract. After 10 min incubation at 37°C , 15 pmol of 50S ribosomal subunits were added. After an additional 10 min incubation, 5 μl of puromycin (5 $\mu\text{g}/\mu\text{l}$) were added. The amount of $fMet-tRNA_f^{Met}$ bound to ribosomes was measured by nitrocellulose filtration, and the amount of formylmethionine reacted with puromycin was measured by ethyl acetate extraction.

B. Binding of $Met-tRNA_f^{Met}$ to 30S ribosomal subunits, effect of 50S subunits, and puromycin sensitivity in the absence (\circ) and presence (\bullet) of initiation factors. Same incubation mixtures as in A, except 3.0 pmol of $Met-tRNA_f^{Met}$ were used instead of $fMet-tRNA_f^{Met}$, and no GTP was added. Incubation periods are indicated with arrows corresponding to the time of the various additions (Petersen et al., 1976a).



The finding that formylation is not essential for the formation of the 30S initiation complex may explain why some *E. coli* mutants can grow in the presence of trimethoprim (which inhibits the enzyme dihydrofolate reductase and thus the formation of the formyl donor ^{10}N -formyltetrahydrofolate (see previous section)).

Nevertheless, the initiator $Met-tRNA_f^{Met}$ has been found to be entirely formylated in almost all *E. coli* cells. What then is the role of the formylation? One explanation is that the N-blocking of the initiator tRNA methionine speeds up the rate of formation of the initiation complex as

suggested by Bretscher & Marcker (1966). A second or additional possible explanation arises from our finding that only the formylated Met-tRNA_f^{Met} binds to 70S ribosomes in the presence of the initiation factors which were discovered after Bretscher's and Marcker's experiments.

Thus, the formyl group is necessary in cases where initiation of translation proceeds via the (untraditional) formation of a 70S initiation complex without a previous involvement of 30S ribosomal subunits. This possibility is investigated in the following section.

A two-state model for the 70S ribosome

In the preceding section it was shown that the 30S ribosomal subunit is apparently not able to distinguish between formylated and unformylated initiator Met-tRNA_f^{Met}, whereas the 70S ribosome discriminates strongly in favour of the formylated species in the presence of initiation factors (which is relevant to the situation *in vivo*) but – also in the absence of initiation factors – it was found that the 70S ribosome can distinguish between the two initiator tRNA species.

The behaviour of the 70S ribosome in this binding reaction has been studied in more detail, as it seems to possess the property to select the N-formylated initiator Met-tRNA_f^{Met} (Petersen *et al.*, 1976b).

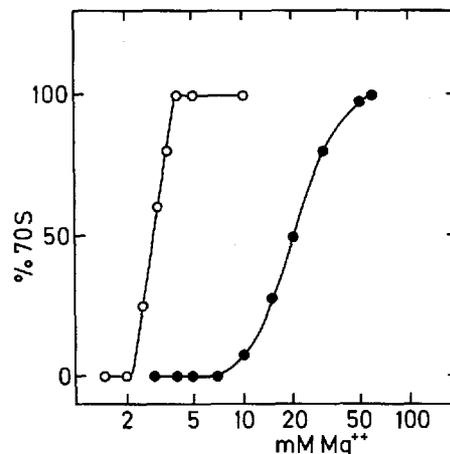
The ability of the 70S ribosome to discriminate between Met-tRNA_f^{Met} and fMet-tRNA_f^{Met} is found to be particularly strong at low

Table 4: Effect of potassium and magnesium ion concentration on the binding of fMet-tRNA_f^{Met} and Met-tRNA_f^{Met} to 70S ribosomes.

Potassium conc. (mM)	Magnesium conc. (mM)	fMet-tRNA _f ^{Met} (pmol)	Met-tRNA _f ^{Met} (pmol)
0	5	0.22	1.29
0	20	1.16	1.33
0	35	1.29	1.11
100	20	0.99	1.20
200	20	0.81	1.23
400	5	0.06	0.18
400	20	0.69	1.21
400	35	1.06	1.02

Incubation for 20 min at 37°C of the following mixture (50 µl): 15 pmoles of 70S ribosomes; 0.11 A₂₆₀ unit of poly(A,G,U); 1 mM GTP; 50 mM Tris-HCl (pH 7.4); 50 mM ammonium chloride; magnesium acetate and potassium chloride as indicated; and 3.0 pmol of fMet-tRNA_f^{Met} or 3.8 pmol of Met-tRNA_f^{Met}. The amount of tRNA bound was analyzed by the Millipore filter assay (Petersen *et al.*, 1976b).

Fig. 10. The equilibrium between 70S ribosomes and 30S + 50S subunits measured by light-scattering at 24°C as a function of magnesium concentration in the absence of potassium (○) and at 400 mM (●) (Petersen et al., 1976b).



magnesium concentration (around 5 mM), whereas it disappears when the Mg^{++} concentration is increased to 20–35 mM (Table 4). It was also found (especially at low Mg^{++} concentrations) that high concentration of potassium (400 mM) strongly inhibits the binding of both initiator tRNA species to the ribosomes. The dissociation curve of ribosomes as a function of magnesium concentration (Fig. 10) shows that the ribosomes are 100% associated at 5 mM Mg^{++} in the absence of potassium (but in the presence of 50 mM ammonium chloride), whereas the ribosomes at 400 mM K^+ and 5 mM Mg^{++} are 100% dissociated. Thus the effect of potassium is in agreement with the results shown in the previous paragraph, neither of the two initiator tRNAs binds to ribosomal subunits in the absence of initiation factors.

$Met-tRNA_f^{Met}$ first bound to 30S subunits can react with puromycin after addition of 50S subunits (Fig. 9). A similar puromycin reaction is not possible for the $Met-tRNA_f^{Met}$ bound directly to the 70S ribosomes in the absence of initiation factors and at low concentrations of K^+ and Mg^{++} (50 mM and 5 mM, respectively). However, when the concentrations of these ions are increased to 400 mM and 35 mM, respectively (almost completely associated ribosomes), the $Met-tRNA_f^{Met}$ binds to a puromycin reactive site on the ribosome (Fig. 11).

The assay for the binding reaction requires filtration on nitrocellulose filters and therefore a reasonably high affinity between the compounds we want to isolate in complexed form, whereas the puromycin reaction takes place on the ribosome and liberates into solution the puromycin covalently bound to the radioactive amino acid, and the reaction product is easily extracted by ethylacetate. Thus, a much weaker binding to the

Table 5: Effect of messenger RNA (poly(A,G,U)) on the binding and puromycin reaction of fMet-tRNA_f^{Met} and Met-tRNA_f^{Met}

	(pmol)	
	- poly(A,G,U)	+ poly(A,G,U)
fMet-tRNA _f ^{Met} binding	0.30	1.05
fMet-puromycin	2.01	1.68
Met-tRNA _f ^{Met} binding	0.11	0.95
Met-puromycin	1.75	1.41

Same incubation mixture as in Table 4 except that GTP was omitted; the amounts of magnesium acetate and of potassium chloride were 35 and 400 mM, respectively. Binding was measured after 20 min at 37°C. For the puromycin reaction, 25 µg of puromycin was added after 20 min at 37°C, and the reaction was continued for 5 min. The reacted amounts were analyzed by the ethyl acetate extraction technique (Petersen *et al.*, 1976b).

ribosome of the Met-tRNA_f^{Met} can be detected by the puromycin reaction compared to the "binding assay". It was therefore interesting to compare the effect of mRNA (synthetic poly(A,G,U)) in the two assays. As shown in Table 2, the binding of Met- and fMet-tRNA_f^{Met} to 30S subunits in the presence of initiation factors depends completely on the presence of a messenger RNA. Table 5 shows that the binding to 70S ribosomes also is messenger-dependent, whereas the puromycin reactivity is found in the total absence of messenger.

It is also shown that the binding of Met-tRNA_f^{Met} as well as fMet-tRNA_f^{Met} is uninfluenced by tetracyclin, an antibiotic which at low con-

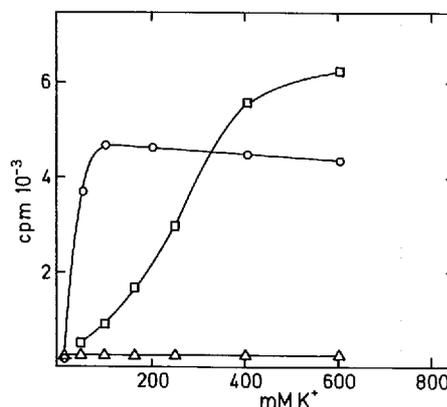


Fig. 11. Puromycin reaction of prebound initiator tRNA as a function of potassium concentration. The incubation mixture contained: 16 pmol of 70S ribosomes, 50 mM Tris-HCl (pH 7.4), 1 mM GTP, 35 mM magnesium acetate and 4.2 pmol of fMet-tRNA_f^{Met} (○), 3.8 pmol of Met-tRNA_f^{Met} (□) or 4.6 pmol of Met-tRNA_m^{Met} (△). The reaction products were analyzed as described in the legend to Figure 9 (Petersen *et al.*, 1976b).

Table 6: Effect of tetracyclin on the binding of fMet-tRNA_f^{Met}, Met-tRNA_f^{Met} and Met-tRNA_m^{Met} to 70S ribosomes

	Translation Factors	tetracyclin	tRNA bound (pmol)
fMet-tRNA _f ^{Met}	none	—	0.55
"	"	+	0.61
"	crude IF	—	1.44
"	"	+	1.47
Met-tRNA _f ^{Met}	none	—	1.96
"	"	+	2.20
"	crude IF	—	0.39
"	"	+	0.52
Met-tRNA _m ^{Met}	none	—	0.42
"	"	+	0.42
"	EF-Tu	—	0.84
"	"	+	0.40

Incubation mixtures were as in Table 4, except for the amount of magnesium acetate which was 5.5 mM and, where indicated: 50 µg of crude initiation factors (IF); 50 µg of elongation factor EF-Tu; 0.1 mM tetracyclin; 2.3 pmol of fMet-tRNA_f^{Met}; 5.2 pmol of Met-tRNA_f^{Met}; or 4.6 pmol of Met-tRNA_m^{Met} (Petersen *et al.*, 1976b).

centrations inhibits the binding of aminoacyl-tRNA to the ribosomal A-site (Table 6).

Parallel experiments were made with the elongator Met-tRNA_m^{Met}. In no case was this tRNA bound in response to initiation factors or showed any reaction with puromycin.

Thus, the 70S ribosome is able to distinguish between the initiator Met-tRNA_f^{Met} and the elongator Met-tRNA_m^{Met}. In the absence of initiation factors, the initiator Met-tRNA_f^{Met} binds very well to 70S ribosomes and can be triggered to react with puromycin. Under these conditions the elongator tRNA_m^{Met} does not bind to the ribosome and reacts under no conditions with puromycin. Furthermore, the results show that 70S ribosomes discriminate strongly between formylated and unformylated initiator Met-tRNA_f^{Met} both in the absence and in the presence of initiation factors. Contrary to the unformylated initiator tRNA, fMet-tRNA_f^{Met} does not bind to 70S ribosomes in the absence of initiation factors. The amount of Met-tRNA_f^{Met} bound in the absence of initiation factors is unaffected by tetracyclin.

This raises the following question:

- 1: Do two different 70S ribosomal sites exist for binding of fMet-tRNA_f^{Met} and Met-tRNA_f^{Met}, respectively, or
- 2: Do two different conformations exist of the ribosomal P-site – the initiator tRNA binding site?

The latter is most likely because: The Met-tRNA_f^{Met} does not bind to the A-site (no effect of tetracyclin) but can be triggered to react with puromycin (the definition of P-site bound aa-tRNA).

Therefore, we propose the existence of the equilibrium:



between a 70S ribosome containing a P-site with inactive peptidyl transferase and a 70S* ribosome with an active peptidyl transferase (or a P*-site).

This also suggests a function of the formylation of the initiator tRNA as fMet-tRNA_f^{Met} only binds to the active P*-site and thus pulls the equilibrium towards the active conformation, whereas Met-tRNA_f^{Met} binds equally well to both conformations and thus does not change the equilibrium. Fig. 12 shows how Met-tRNA_f^{Met} (M) and fMet-tRNA_f^{Met} (FM) are supposed to bind to the two conformations at different conditions.

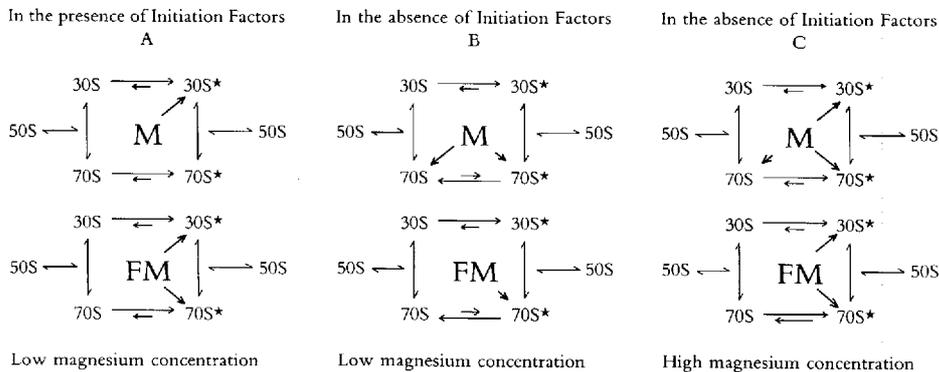


Fig. 12. A schematical presentation of how the position of the 30S and 70S ribosomal initiator tRNA binding site conformational equilibrium is supposed to be at low magnesium concentration (5 mM) (A and B), and high magnesium concentration (35 mM) (C) in the presence (A) and absence (B and C) of initiation factors. The length of the arrows from each initiator tRNA species indicates the relative affinity for the different ribosomal conformations, and the sum of arrows from one tRNA indicates the total binding at the respective conditions. M: Met-tRNA_f^{Met}, FM: fMet-tRNA_f^{Met} (Petersen, 1980).

TP9 – a puromycin resistant E. coli mutant

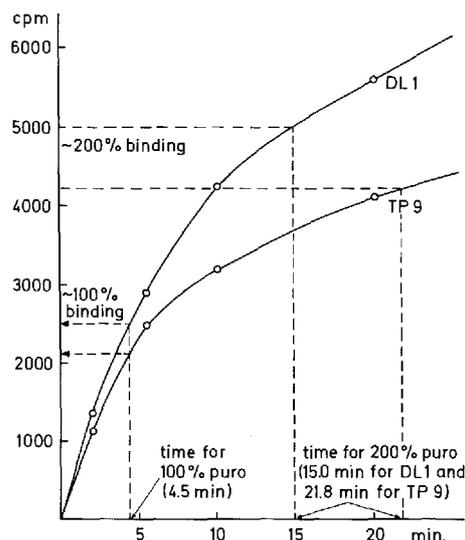
As a consequence of the proposed hypothesis that the ribosomal P-site exists in two conformations, one would expect the equilibrium between these two conformations to be sensitive to mutations in the ribosomal components involved in the tRNA binding area. We know that puromycin binds to the ribosome in the region between the A-site and the P-site (Cooperman, 1980), and it would therefore be expected that bacterial strains permeable for but resistant to puromycin would be affected in the $70S \rightleftharpoons 70S^*$ equilibrium.

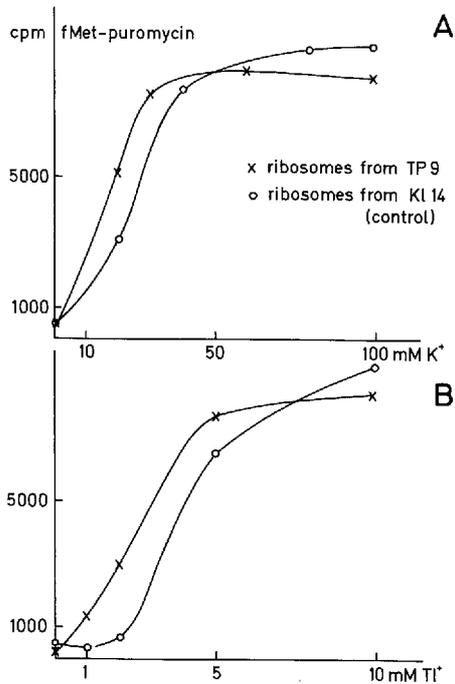
From the parental strain DL1 (a puromycin sensitive strain of *E. coli* K12), Dr. Antoine Danchin, Institut Pasteur, Paris, isolated a mutant (TP9) which was able to grow in broth medium supplied with the C-1 metabolites plus puromycin and trimethoprim.

In collaboration with Dr. Leif Isaksson at the Wallenberg laboratory, Uppsala, ribosomes isolated from TP9 were analyzed by two-dimensional polyacrylamide gel electrophoresis, and we found displaced positions of the ribosomal proteins S7, S20 (= L26) and L27. S7 is located at the tetracyclin binding site (A-site), and L27 is part of the peptidyl transferase centre and thus both the ribosomal A- and P-sites (Ofengand *et al.*, 1984). S20 is situated at the "head" of the 30S subunit and most probably at the interface between the two subunits (Stöffler *et al.*, 1979).

Ribosomes isolated from TP9 and DL1 were used in *in vitro* experiments where the kinetics of binding and puromycin reaction of Met-

Fig. 13. Kinetics of the puromycin reaction of Met-tRNA^{Met} bound to 70S ribosomes from wild type (DL1) and puromycin resistant (TP9) *E. coli* cells. The curves show the amount of Met-puromycin formed as a function of time after 20 min binding incubation at 37°C in 50 mM Tris-HCl (pH 7.5), 35 mM magnesium chloride, 150 mM potassium chloride and 64 pmol of Met-tRNA^{Met}. The radioactivity corresponding to the amount of Met-tRNA^{Met} bound to the ribosomes after 20 min is indicated (100%). Dotted lines indicate the time required for a puromycin reaction of twice the amount of Met-tRNA bound after 20 min (15.0 min for wild type and 21.8 min for TP9 ribosomes) (Petersen, 1980).





A Fig. 14. Puromycin reaction of fMet-tRNA^{Met} bound to wild type (K114) and TP9 ribosomes as a function of (A) potassium and (B) thallium concentration. Incubation conditions as described in the legend to Figure 13. Ethyl acetate extraction after 5 min puromycin reaction (Petersen, 1980).

tRNA_f^{Met} were measured. No difference was found in the kinetics of the binding reaction, whereas a difference was observed in the puromycin reaction at prolonged time of incubation. As shown in Figure 13, the time needed to obtain a puromycin reaction corresponding to the amount of Met-tRNA_f^{Met} bound is the same for the two types of ribosomes. A higher amount of puromycin reaction can be obtained if the incubation is continued, but in that case, the reaction proceeds much more slowly with the mutant ribosomes as compared to the control ribosomes.

This result may indicate that the mutation leading to puromycin resistance affects the ribosomal site of translocation in some so far unknown way.

The binding of fMet-tRNA_f^{Met} to the wild type and mutant ribosomes was studied as a function of K⁺ concentration (Figure 14.A). One sees the previously found sigmoidal shaped curve for the control ribosomes – indicating the double function of potassium ions in the 70S ⇌ 70S* equilibrium and in the puromycin reaction, whereas the corresponding experiment with TP9 ribosomes gives a hyperbolic shaped binding curve. This indicates that with TP9 ribosomes, one of the two functions

of K^+ is suppressed – possibly the effect on the $70S \rightleftharpoons 70S^*$ equilibrium – as this may be strongly displaced toward the active form in the mutant.

As potassium ions may have a specific catalytic site at the peptidyl transferase center and not merely be required to obtain a certain ionic environment for the protein biosynthesis, we did similar experiments of binding and puromycin reaction replacing the potassium with varying concentrations of thallium, which is similar to K^+ with respect to stereochemical properties. Fig. 14B shows the results. To obtain similar amounts of fMet-puromycin, 10 times lower concentrations of Tl^+ as compared to K^+ is needed. This is a further indication of a specific ribosomal site for K^+ (Petersen, 1980).

Hypothesis on the role of the formyl group of fMet-tRNA_f^{Met} in the translation of polycistronic messenger RNAs

Based on the *in vitro* experiments, a general function of the formyl group of fMet-tRNA_f^{Met} in initiation of translation has been proposed (Petersen *et al.*, 1976b). In prokaryotes, the traditional view of the mechanism of initiation involved the formation of a 30S initiation complex. As shown in Table 2, the 30S ribosomal subunit (even in the presence of initiation factors) cannot discriminate between formylated and unformylated initiator tRNA, whereas in the presence of initiation factors, only the formylated species binds to 70S ribosomes (Table 3). Now, recall the situation in eukaryotic cells: 1) no formylation of the initiator Met-tRNA exists, and 2) no polycistronic mRNAs are known. Therefore, no initiation is possible at internal initiation sites in eukaryotic mRNAs.

In prokaryotes, the mRNAs commonly are polycistronic and one can imagine that the 5'-end proximal initiation site always (as in eukaryotes) forms an initiation complex with 30S ribosomal subunits. This would not require formylation of initiator tRNA (although the rate of initiation is increased by the formylation which is in agreement with the fact that prokaryotic translation is 9-12 times faster than eukaryotic translation).

If the ribosome translating a polycistronic messenger does not dissociate into subunits after terminating the translation of the first cistron but as a 70S particle migrates through the intercistronic region of the mRNA (which varies considerably in length in different mRNAs), this 70S ribosome will be directly involved in the initiation complex formation at the next initiator codon. As only the formylated initiator tRNA binds to the

Table 7: Effect of trimethoprim, chloramphenicol and kasugamycin on the coordinate synthesis of β -galactosidase (Gz) and thiogalactoside acetyltransferase (Ac)

Strain	Addition ($\mu\text{g/ml}$)		Ratio (Gz/Ac)
CP78	None		1
	Trimethoprim	0.75	1.5
	Trimethoprim	1.5	2.4
	Chloramphenicol	1.5	1.06
	Chloramphenicol	2	1.14
	Kasugamycin	40	0.92
CP781a	None		1
	Trimethoprim	0.75	1.15
	Trimethoprim	1.5	1.2
	Chloramphenicol	1.5	1.1
	Chloramphenicol	2	1.2
	Kasugamycin	20	0.96

(Petersen *et al.*, 1978).

70S ribosome in the presence of initiation factors, the formylation is an absolute requirement in this situation.

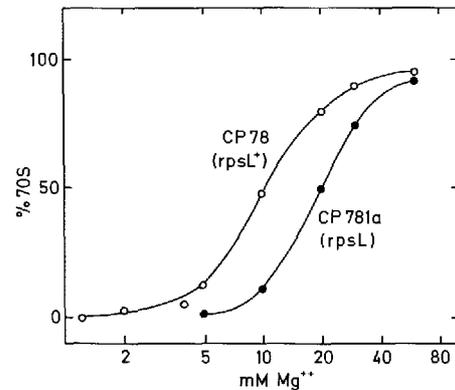
From this hypothesis, one can predict that formylation of initiator tRNA methionine can be related to the polarity in the translation of polycistronic mRNAs in such a way that a decrease in the level of formylation under otherwise normal conditions will result in an increased polarity, 5'-end proximal cistrons being translated relatively more efficiently compared to 5'-end distal cistrons.

In vivo polarity in lactose operon expression, role of formylation of initiator tRNA methionine

It has been observed that *E. coli* mutants which were able to grow in the presence of low levels of trimethoprim – an inhibitor of formylation, see Figure 6 – contained an increased proportion of ribosomal subunits to 70S ribosomes as compared to wild type cells (Harvey, 1973). This is an indication that the cell can overcome the lack of formylation by increasing the proportion of 30S to 70S ribosomes, which is in agreement with the results that 30S and not 70S ribosomes are able to use unformylated Met-tRNA_f^{Met} as initiator tRNA.

It is also expected from our hypothesis that depression of formylation

Fig. 15. Light-scattering measurement of the association of ribosomal subunits as a function of magnesium concentration at 150 mM potassium chloride and at 37°C. Ribosomes from strain CP78 (*rpsL*⁺) (○) and from strain CP781a (*rpsL*) (●) of *E. coli* (Petersen et al., 1978).



in normal growing cells will affect the ratio between the efficiency of translation of the 5'-end proximal and distal cistrons of polycistronic mRNAs in such a way that lack of formylation would favour the synthesis of the first protein rather than later proteins from the operon.

This is tested by measuring the *in vivo* translation of the first and the last cistron of the lactose operon in *E. coli* – β -galactosidase and thiogalactoside acetyltransferase, respectively (Petersen *et al.*, 1978). We have studied the effect of inhibition of formylation of initiator tRNA methionine on the relative synthesis of these two proteins in wild type cells and in *E. coli* mutants resistant to streptomycin and trimethoprim.

The presence of trimethoprim in the growth medium has a significant effect on the differential translation of the first and last cistrons of the lactose operon in *E. coli* wild type cells. This inhibitor of formylation results in a two-fold increase in the ratio: β -galactosidase/thiogalactoside acetyltransferase synthesized in wild type cells, whereas no significant effect was found in streptomycin resistant strains (Table 7). Control experiments with other antibiotic inhibitors (kasugamycin and chloramphenicol – which are inhibitors of initiation of translation without affecting the formylation of initiator tRNA methionine) showed that these compounds had no effect on the relative expression of the different cistrons in any of the tested strains.

Ribosomes were extracted from all tested strains and dissociation curves were measured. This showed that at conditions of pH, temperature and concentrations of mono- and divalent cations comparable to those found *in vivo*, the ribosomes from the mutant cells were significantly more dissociated than those from wild type cells (Fig. 15).

These results are consistent with the hypothesis that a significant part

of the ribosomes may pass through the intercistronic region from the previous termination step to the following initiation step without dissociating off the messenger. This could be the case when an unlimited amount of fMet-tRNA_f^{Met} is available for the initiation, whereas the ribosome must dissociate and use a 30S subunit for the following initiation at low levels of formylation.

Our suggested mechanism also explains why, in eukaryotic cytoplasm, no formylation of initiator tRNA methionine is needed – as no cellular polycistronic mRNAs apparently exist.

Structural elements in initiator tRNA involved in specific interactions with proteins

As a consequence of the results discussed above, we have been interested in further investigation of the question: What are the specific structural elements in tRNA_f^{Met} which are involved in and thus responsible for the different specific interactions with other macromolecules during the initiation process as shown schematically in Fig. 8? Recent developments in the methodology of nucleic acids research has made such investigations of molecular details possible.

In the following paragraphs, we will look a little more closely at the regions within the initiator tRNA_f^{Met} which are implicated in these interactions. The effect of formylation is studied, and the results are compared to experiments with the elongator tRNA_m^{Met} in the cases where the two tRNAs are involved in similar macromolecular interactions.

Two methods have been employed in these investigations: The footprinting method, in which ribonucleases are used for cutting the tRNA molecule at specific sites. The resulting RNA fragments are analyzed by polyacrylamide gel electrophoresis, which makes it possible to determine the exact cutting positions in the nucleotide sequence. Results from the treatment of free tRNA are compared to those from tRNA complexed to a protein molecule. In the case where different ribonucleases cut a particular position less efficiently in complexed tRNA as compared to free tRNA, it is concluded that this position is protected by the protein. The technical details of this method has been described recently (Petersen *et al.*, 1984c).

The second method is based on the lability of the aminoacyl-ester bond in aminoacyl-tRNAs. The reaction of hydrolysis of this linkage is

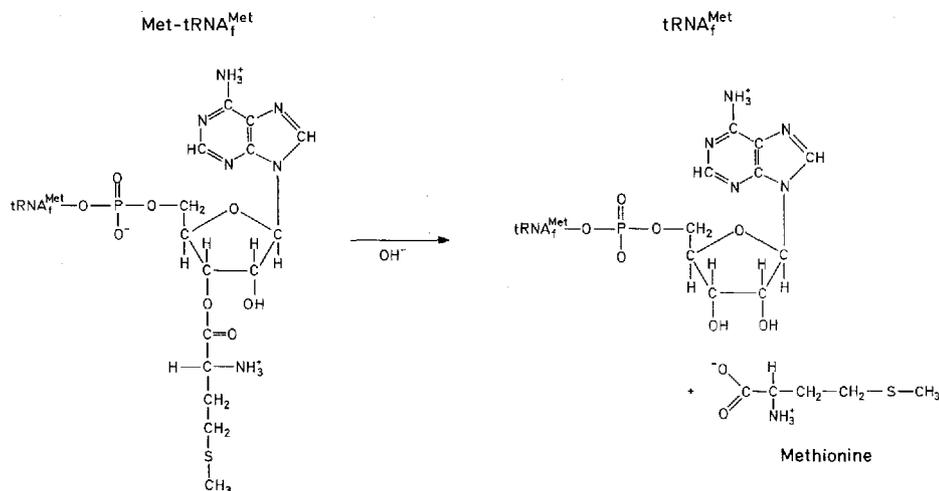


Fig. 16. Reaction of spontaneous hydrolysis of the aminoacyl ester bond in $\text{Met-tRNA}_f^{\text{Met}}$.

shown in Fig. 16. At neutral pH, the ester bond is hydrolyzed spontaneously at a rate which among other factors depends on the temperature. This method has been applied successfully in the study of interactions between elongation factor EF-Tu, GTP and different aminoacyl-tRNAs (Pingoud *et al.*, 1977). The rate of hydrolysis of $\text{fMet-tRNA}_f^{\text{Met}}$ or $\text{Met-tRNA}_f^{\text{Met}}$ is measured in the absence or presence of proteins. In the case where the half-life of the ester bond is increased by the addition of a protein, it is concluded that an interaction takes place between the two macromolecules and in particular that a closer contact must take place at the amino acid attachment site.

In the following, results are described from such studies on interactions with methionyl-tRNA synthetase, initiation factor IF2, elongation factor EF-Tu and the 70S ribosome.

Interaction with methionyl-tRNA synthetase (MetRS)

Some characteristic features of the structure of the initiator $\text{tRNA}_f^{\text{Met}}$ as compared to that of the elongator $\text{tRNA}_m^{\text{Met}}$ can be studied by comparing the footprinting results from complexes between methionyl-tRNA synthetase and the two tRNAs. As described in an earlier paragraph, the enzyme aminoacylates both tRNAs with methionine in the reactions 1 and 2 shown in Figs. 8 and 23. Although, as shown, the tRNA enzyme interaction *in vivo* takes place after the binding of a methionyl-adenylate

to the enzyme (which may change the conformation of the enzyme), a stable complex between $\text{tRNA}_f^{\text{Met}}$ or $\text{tRNA}_m^{\text{Met}}$ and MetRS can be formed and isolated *in vitro* in the absence of other ligands (complex I in Fig. 8 and complex I in Fig. 23).

Results from experiments using T1, T2 and cobra venom RNases on the binary complex enzyme:tRNA with each of the two tRNAs are shown in Figs. 17 and 18 (Petersen *et al.*, 1984a). Regions which are protected by MetRS are shown in black, whereas locations which seem more accessible for nucleases in the complexed tRNAs are shown hatched.

The 3'-side of the anticodon loop is markedly protected by MetRS in both tRNAs. Differences are seen in the 5'-side of the anticodon stem, which is protected in $\text{tRNA}_f^{\text{Met}}$, but cut more intensely in $\text{tRNA}_m^{\text{Met}}$ when complexed to MetRS. Alternatively, the D-loop and the extra-loop are

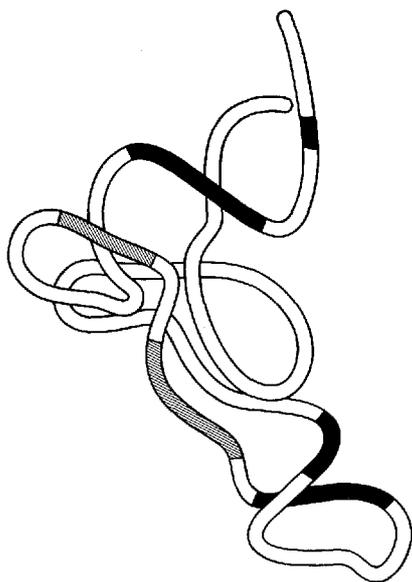


Fig. 17. Footprinting results from $\text{tRNA}_f^{\text{Met}}$ complexed to dimeric methionyl-tRNA synthetase are transferred to a tertiary structure model of tRNA^{Phe} (see Fig. 3). Regions within $\text{tRNA}_f^{\text{Met}}$ which are protected (■) or cut more intensely by ribonucleases (▨) when complexed to MetRS are indicated (Petersen *et al.*, 1984a).

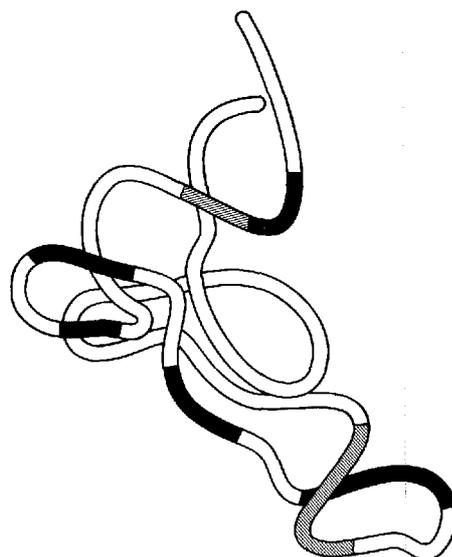


Fig. 18. Footprinting results from $\text{tRNA}_m^{\text{Met}}$ complexed to dimeric methionyl-tRNA synthetase are transferred to a tertiary structure model of tRNA^{Phe} as described in the legend to Fig. 17 (Petersen *et al.*, 1984a).

protected in complexed tRNA_m^{Met}, whereas both these regions are cut more intensely by RNases in tRNA_f^{Met} as a result of complex formation.

It is known that cobra venom ribonuclease cuts tRNA at many positions in the acceptor region (Boutorin *et al.*, 1981). MetRS shows only weak protection in the amino acid region. This supports the idea that (at least in the absence of the other substrates of the aminoacylation reaction), the acceptor region of tRNA is not strongly bound at the surface of the enzyme (Jacques & Blanquet, 1977).

Recent data indicate that also in the presence of methionine and ATP, no strong protection against cobra venom RNase digestion in the acceptor region takes place (H. U. Petersen, G. E. Siboska & S. Blanquet, unpublished). This result is in good agreement with the effect of MetRS on the spontaneous hydrolysis of the aminoacyl ester bond in fMet-tRNA_f^{Met} in the presence of the initiation factor IF2, which is discussed in the next paragraph.

The interactions with the synthetase seem to be similar in the extreme parts (amino acid attachment site and anticodon region) of the tRNAs, whereas only tRNA_m^{Met} seems to bind to the protein in the central part of the molecule. Thus, it can be concluded that the three-dimensional structures of the two molecules have differences in the region around the extra loop. In later paragraphs, it will be seen that the extra loop probably plays an important role for the specificity of the initiator tRNA.

Interaction with the initiation factor IF2

The initiation factor IF2 exists in 2 forms: IF2 α (M_r 97300) and IF2 β (M_r 79700), which are coded for by the same gene (Plumbridge *et al.*, 1985). The DNA sequence of this gene has recently been determined in the laboratory of Dr. M. Grunberg-Manago (Sacerdot *et al.*, 1984). As the available amounts of pure initiation factor protein has been used for functional studies and no attempt to crystallize the protein has been done, no information has been obtained on the higher order structure of these proteins. However, using small angle neutron scattering (at the equipment D11 of the Institut Laue-Langevin in Grenoble, France) we have determined the radius of gyration, R_g, of IF2 to be 45–48Å in solution, which indicates that the protein has a rather elongated shape (H. U. Petersen, M. Grunberg-Manago and B. Jacrot, unpublished).

Whilst it is well established that the protein chain elongation factor EF-Tu functions as an aminoacyl-tRNA carrier protein in a tertiary complex: EF-Tu:GTP:aa-tRNA during the elongation step of the prokaryotic translation, the question whether a similar complex is formed be-

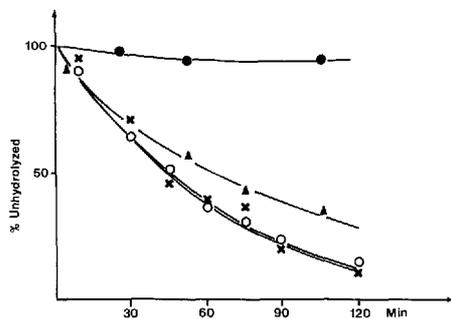


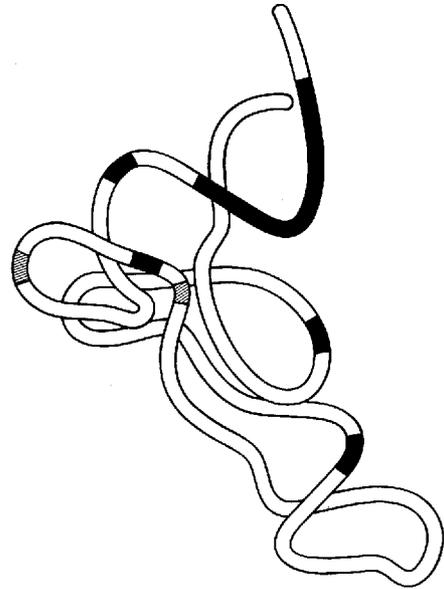
Fig. 19. Kinetics of the non-enzymatic hydrolysis of the aminoacyl ester bond of fMet-tRNA_f^{Met} at 37°C in the absence (○) and presence (●) of initiation factor IF2 at 8 times molar excess, MetRS at 16 times molar excess (x) or both IF2 and MetRS at 8 and 16 times molar excess, respectively, (▲) (Petersen et al., 1984c).

tween the initiation factor IF2, GTP and fMet-tRNA_f^{Met} during chain initiation (complex II in Fig. 8) has been the subject of intense debate for more than a decade. Although many different chemical and physical methods have been applied in such investigations, a complex of native IF2, fMet-tRNA_f^{Met} and GTP has never been isolated (Petersen *et al.*, 1979).

Previous studies have shown that IF2 invariably interacts with fMet-tRNA_f^{Met}. However, the extent of interaction with unformylated Met-tRNA_f^{Met} varied from no detectable interaction to almost the same level as for the formylated species. Although it is clear that GTP hydrolysis is necessary for the formation of the functional 70S ribosomal initiation complex, no experiments have shown that GTP is required for IF2 to interact with the initiator tRNA. In addition, the ionic requirements for the formation of a binary complex IF2:fMet-tRNA_f^{Met} has varied considerably. A general feature of all earlier experiments is the attempt to isolate a macromolecular complex, and such a complex may dissociate during the preparation, whereas the methods employed here do not require the isolation of a complex.

Fig. 19 shows the rate of hydrolysis of the aminoacyl-esterbond of fMet-tRNA_f^{Met}, free and in the presence of MetRS, IF2 or both proteins (Petersen *et al.*, 1984c). This shows that MetRS has no effect itself on the rate of hydrolysis. The lack of protective effect is an indication that MetRS does not bind strongly at the aminoacyl-linkage. It also shows that MetRS (in the absence of AMP) does not catalyze the de-aminoacylation. In the presence of IF2, complete protection is observed. When increasing amounts of MetRS are added in the presence of sufficient amount of IF2 to obtain such complete protection, the effect of IF2 disappears. This is taken as an indication that MetRS and IF2 have overlapping binding sites on the tRNA molecule and, thus, in this experi-

Fig. 20. Footprinting results from fMet-tRNA^{Met} in the presence of initiation factor IF2 are transferred to a tertiary structure model of tRNA^{Phe} as described in the legend to Fig. 17 (Petersen et al., 1984c).



ment, compete for the binding to this site. However, it is also clear that the binding site of MetRS is not the 3'-terminal part of fMet-tRNA^{Met}, which, on the other hand, seems to be one of the important binding sites for IF2.

Similar experiments were done using unformylated initiator Met-tRNA^{Met} or elongator Met-tRNA^{Met} (Petersen *et al.*, 1979). In both cases, no effect was observed on the rate of hydrolysis when increasing amounts of IF2 were added. Although this method does not exclude the possibility that the protein binds to the tRNA, it is clear that an interaction similar to the one found with fMet-tRNA^{Met} does not take place.

The binding site of IF2 on fMet-tRNA^{Met} has been further investigated by footprinting experiments (Petersen *et al.*, 1981). Fig. 20 shows the regions which are protected or cut more intensely in fMet-tRNA^{Met} complexed to IF2. In this case, only the double strand specific ribonuclease isolated from the venom of *Naja oxiana* has been used to digest the tRNA.

The protected regions include the 3'-end, both sides of the T-stem, the anticodon stem (in particular the 5'-side) and the D-stem. It seems that the protected regions are mainly located at the extreme parts of the L-shaped tRNA molecule, and no protection is found in the extra loop. This result is very similar to the one found with MetRS except that IF2 seems to be in closer contact with the 3'-end of the tRNA.

Footprinting experiments were also done with unformylated initiator tRNA. However, IF2 had no effect on the results obtained with free Met-tRNA_f^{Met}. This is a further indication that no binding takes place between IF2 and Met-tRNA_f^{Met} and supports the idea that formylation adds a signal to the initiator tRNA for the specific interaction with IF2 in the absence of ribosomes.

Interaction with the 70S ribosome

As shown in reaction 7 of Fig. 8 and as discussed in detail in previous paragraphs, fMet-tRNA_f^{Met} can bind non-enzymatically to the ribosomal P-site at 15-20 mM magnesium ion concentration. The complex described here was formed in this way in the absence of initiation factors, using a poly(A,G,U) RNA chain as a messenger (Petersen *et al.*, 1984c). In order to ensure that no unbound tRNA is present in the footprinting study, the complex was isolated on a Sepharose 6B column prior to enzymic digestion. Fig. 21 shows the regions in the tRNA_f^{Met}, which are protected against ribonuclease (RNases T1 and T2) digestion in the initiation complex in the absence of initiation factors. These are seen to be located in the aminoacyl-stem, the variable loop and in the part of the anticodon stem close to the extra loop. When compared to the footprinting results with MetRS and IF2, this is the first time we observe a strong protection of tRNA_f^{Met} in the extra loop. The involvement of the extra

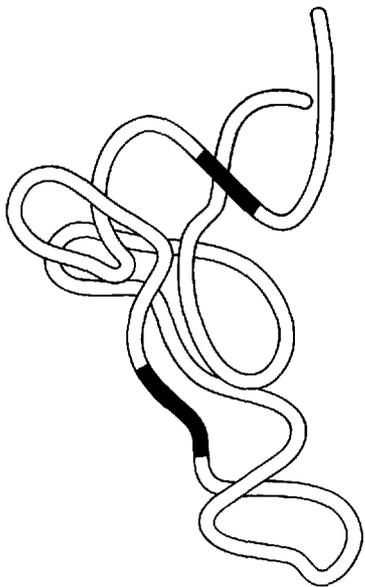


Fig. 21. Footprinting results from fMet-tRNA_f^{Met} bound non-enzymatically to 70S ribosomes are transferred to a tertiary structure model of tRNA^{Phe} as described in the legend to Fig. 17 (Petersen *et al.*, 1984c).

loop in tRNA interaction with the 70S ribosome has also been observed for the A-site binding of tRNA^{Phe} (Petersen *et al.*, 1984a), whereas other regions within the two tRNA molecules seem to have different degrees of involvement in A-site and P-site interactions (T. Jørgensen, personal communication). It must be noted that all the protected regions are found located at one side of the tRNA_f^{Met} molecule, the side exposing the extra loop. Although it is too early to make a firm conclusion, this indicates the contact-face of the initiator tRNA when bound at the 70S ribosomal P-site.

Interaction with elongation factor EF-Tu

During the elongation steps of protein biosynthesis, aminoacyl-tRNAs form a ternary complex with the elongation factor EF-Tu and GTP. Elongation factor EF-Tu is an acidic protein consisting of 393 amino acids. The amino acid sequence has been determined in Aarhus (Jones *et al.*, 1980), where also a three-dimensional structure model of the molecule at 2.9 Å resolution has been obtained by X-ray crystallography (Morikawa *et al.*, 1978; Rubin *et al.*, 1981). Using the footprinting method, we have been investigating the interaction between this protein and different aminoacyl-tRNAs – in particular the elongator tRNA_m^{Met} (Boutorin *et al.*, 1981; Wikman *et al.*, 1982). A summary of these results is seen in Fig. 22. In this case as well, all protected sites are found at one side of the tRNA_m^{Met} molecule. Again, the extra loop seems involved in the interaction together with parts of the T-arm and the region near the amino acid attachment site.

As a comparison with tRNA_f^{Met} (Fig. 8), the interactions involving tRNA_m^{Met} prior to the binding at the ribosomal A-site in protein biosynthesis are shown schematically in Fig. 23. The aminoacylation is catalyzed by MetRS (Reactions 1 and 2), and the synthetase is probably displaced by EF-Tu:GTP during the formation of the ternary complex (Reaction 3). Subsequently, the Met-tRNA_m^{Met} is bound at the 70S ribosomal A-site carried by EF-Tu (Reaction 4).

Until recently, it was believed that only tRNA_m^{Met} and not tRNA_f^{Met} could take part in these two last reactions (reactions 3 and 4 in Fig. 23) (Ofengand, 1977). However, new experiments based on the protection by EF-Tu against pancreatic RNase digestion of the (unformylated) initiator Met-tRNA_f^{Met} revealed that a ternary complex Met-tRNA_f^{Met}:EF-Tu:GTP can actually be formed (Tanada *et al.*, 1982). This is shown as reaction 8 in Fig. 8. We have confirmed this observation by isolating a ternary complex Met-tRNA_f^{Met}:EF-Tu:GTP. In experiments similar to

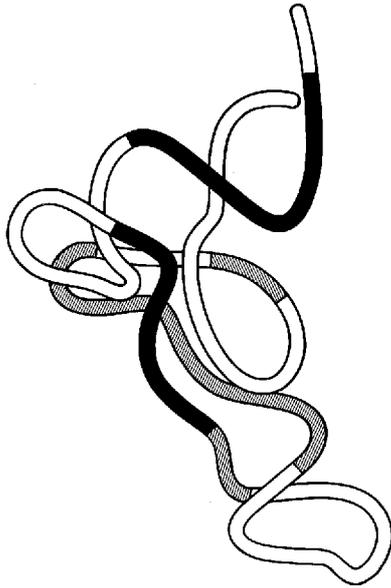


Fig. 22. Footprinting results from Met-tRNA_m^{Met} complexed to the elongation factor EF-Tu and GTP are transferred to a tertiary structure model of tRNA^{Phe} as described in the legend to Fig. 17 (Wikman et al., 1982).

those shown in Fig. 19, we obtained a strong protection against hydrolysis of Met-tRNA_f^{Met} by addition of EF-Tu:GTP, whereas no effect is found on the rate of hydrolysis of fMet-tRNA_f^{Met} (P. Kamp Hansen, H. U. Petersen, J. W. B. Hershey & B. F. C. Clark, unpublished).

Thus, EF-Tu:GTP in bacteria may well discriminate against only the species fMet-tRNA_f^{Met}. The formylation could help play a role as security against the Met-tRNA_f^{Met} acting as an elongator and translating incor-

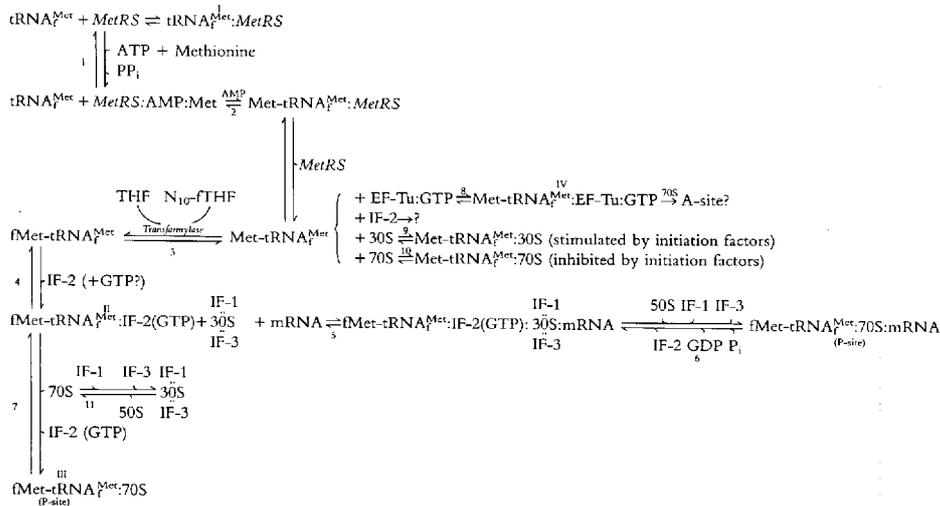


Fig. 23. Reactions involving the elongator tRNA_m^{Met} prior to the binding at the 70S ribosomal A-site.

rectly at GUG and UUG, which the initiator species theoretically can do during initiation (Clark & Marcker, 1966b). What happens during eukaryotic protein biosynthesis in this connection is rather unclear, because the initiator species itself is not formylated and exists as Met-tRNA_i^{Met}. However, the whole selection of the initiator tRNA is more restricted in eukaryotes in the sense that more protein components are involved (Hershey *et al.*, 1984). Of course, during bacterial protein initiation, as discussed in previous sections, initiation factors play a significant role in placing the initiator tRNA in the correct P-site for initiation. More detailed information about this interaction of the initiator tRNA were obtained by footprinting studies on the complex similar to those described above for elongator tRNA. Our results (Clark *et al.*, 1984) are shown in Fig. 24. Although the studies have not been so extensive, we observed a similar general pattern of protection by EF-Tu:GTP against nuclease cutting of the initiator tRNA as for the elongator tRNA_m^{Met} (Fig. 22) but with some small differences.

Again, we see protection in the aa-stem, T-stem and extra loop. In addition, the protection runs into the T-loop on both sides where we saw no changes for the elongator tRNA.

We are not able to decide at this time whether this should be interpreted in terms of extra covering by EF-Tu:GTP of the tRNA or steric hindrance due to local conformational changes.

What appears to be a reasonable conclusion is that bacterial Met-

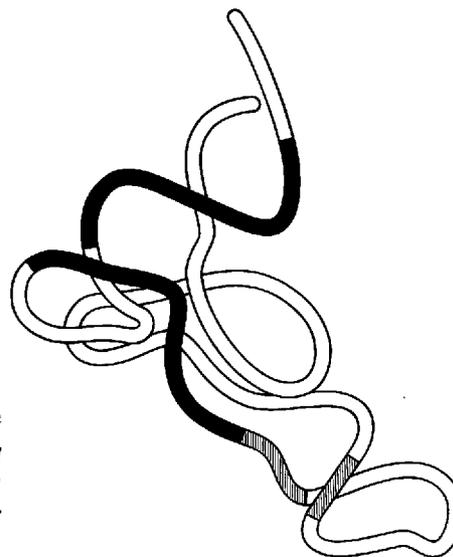


Fig. 24. Footprinting results from Met-tRNA_i^{Met} complexed to the elongation factor EF-Tu and GTP are transferred to a tertiary structure model of tRNA^{Phe} as described in the legend to Fig. 17 (Clark *et al.*, 1984).

$tRNA_f^{Met}$, the initiator tRNA in the unformylated state, can indeed form a ternary complex with EF-Tu:GTP. This ternary complex also appears almost normal in the disposition of the aa-tRNA on the EF-Tu:GTP, so it should be able to bind to the ribosomal A-site if its formation is possible *in vivo*. We are presently investigating whether the small differences in the T-loop binding are sufficient for making a fit poor enough in the A-site to be selected against.

Conclusion

In summary this paper has described work on the function of initiator tRNA during the steps of protein biosynthesis initiation. We have suggested a role for the specific formylation of prokaryotic initiator tRNA methionine in the translation of polycistronic messenger RNAs. The

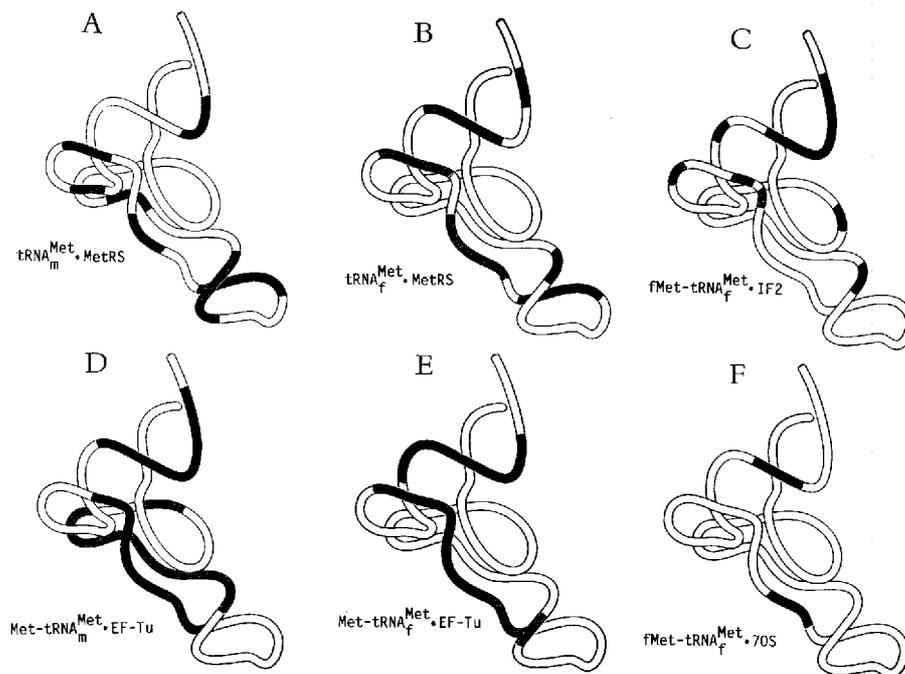


Fig. 25. A summary of the footprinting results shown in Figures 17, 18, 20-22, and 24. Regions protected in complexed tRNA are shown in red and regions cut more intensely are green. (A): $tRNA_m^{Met}$: Met-RS; (B): $tRNA_f^{Met}$: Met-RS; (C): $fMet-tRNA_f^{Met}$: IF2 (D): $Met-tRNA_m^{Met}$: EF-Tu; (E): $Met-tRNA_f^{Met}$: EF-Tu; (F): $fMet-tRNA_f^{Met}$: 70S.

regions within this tRNA which are involved in the interactions with proteins during the initiation process have been studied and the results are compared in figure 25. Although a firm conclusion about the exact sites of interactions with proteins requires further investigations – and ultimately the crystallization and X-ray diffraction structural determination of the tRNA within the complex – our results indicate some similarities and differences between the tRNA_f^{Met} interactions with MetRS, IF2, EF-Tu and 70S ribosomes and also between the interactions with MetRS or EF-Tu of tRNA_f^{Met} and the elongator tRNA_m^{Met}. The regions within the two tRNAs which are protected by proteins are shown in red and the regions which are cut more intensely in complexed tRNA are green.

ACKNOWLEDGEMENT. The work described is a review of research carried out partly in the laboratory of dr. M. Grunberg-Manago at Institut de Biologie Physico-chimique in Paris in collaboration with dr. A. Danchin. The structural studies took place at the Department of Biostructural Chemistry at Aarhus University in collaboration with professor B. F. C. Clark and the personnel of the Enzymelaboratory. A fruitful collaboration also includes drs. R. H. Buckingham, Institut de Biologie Physico-Chimique, Paris, S. Blanquet, Ecole Polytechnique, Palaiseau and J.-P. Ebel, Institut de Biologie Moléculaire et Cellulaire, Strasbourg.

I would like to express my warmest gratitude to my coworkers who have been involved in the described work. Ms. L. Heilesen is thanked for help with the preparation of this manuscript and Mr. H. O. Petersen, A. Lindahl and O. Jensen for technical and photographic help in the preparation of the figures.

In addition to the support from the Niels Bohr Fellowship Committee, the project has received grants from The Danish Natural Science Research Council, The European Molecular Biology Organization (EMBO), The Carlsberg Foundation and The Ministry of Foreign Affairs, The French Government. The support is much appreciated.

References

- Baumstark, B. R., Spremulli, L. L., RayBhanaary, U. L. & Brown, G. M. (1977) *J. Bacteriol.* 129, 457-471.
- Bishop, J., Leaky, J. & Schwett, R. (1960) *Proc. Nat. Acad. Sci. USA* 46, 1030-1038.
- Björk, G. and Isaksson, L. (1970). *J. Mol Biol.* 51, 85-100.
- Blanquet, S., Dessen, P. & Iwatsubo, A. (1976) *J. Mol. Biol.* 103, 765-784.
- Blanquet, S., Iwatsubo, M. & Waller, J.-P. (1973) *Eur. J. Biochem.* 36, 213-226.
- Boutorin, A. S., Clark, B. F. C., Ebel, J. P., Kruse, T. A., Petersen, H. U., Remy, P. & Vassilenko, S. (1981) *J. Mol. Biol.* 152, 593-608.
- Bretscher, M. S. & Marcker, K. A. (1966) *Nature* 211, 380-384.
- Clark, B. F. C., la Cour, T. F. M., Nielsen, K. M., Nyborg, J., Petersen, H. U., Siboska, G. E. & Wikman, F. P. (1984) in "Gene Expression. The Translational Step and its Control" (eds. B. F. C. Clark & H. U. Petersen) Munksgaard, Copenhagen, pp. 127-148.
- Clark, B. F. C. & Marcker, K. A. (1966a) *Nature* 211, 378-380.
- Clark, B. F. C. & Marcker, K. A. (1966b) *J. Mol. Biol.* 17, 394-406.
- Cooperman, B. S. (1980) in "Ribosomes, Structure, Function and Genetics" (eds. G. Chambliss, G. R. Craven, J. Davis, K. Davis, L. Kahan & M. Nomura) University Park Press, Baltimore, pp. 531-554.
- Cory, S., Marcker, K. A., Dube, S. K. & Clark, B. F. C. (1968) *Nature* 220, 1039-1040.
- Danchin, A. (1973) *FEBS Lett.* 34, 327-332.
- Delk, A. S. & Rabinowitz, J. C. (1975) *proc. Nat. Acad. Sci. USA* 72, 528-530.
- Dessen, P., Blanquet, S., Zaccai, G. & Jacrot, B. (1978) *J. Mol. Biol.* 126, 293-313.
- Dessen, P., Fayat, G., Zaccai, G. & Blanquet, S. (1982) *J. Mol. Biol.* 154, 603-613.
- Dickerman, H. W., Steers, E., Redfield, B. G. & Weissbach, H. (1967) *J. Biol. Chem.* 242, 1522-1525.
- Drews, J., Grasmuk, H. & Unger, F. M. (1973) *Biochem. Biophys. Res. Comm.* 51, 804-812.
- Dube, S. K., Marcker, K. A., Clark, B. F. C. & Cory, S. (1968) *Nature* 218, 232-233.
- Eisenstadt, J. & Lengyel, P. (1966) *Science* 154, 524-527.
- Galper, J. B. & Darnell, J. E. (1969) *Biochem. Biophys. Res. Comm.* 34, 205.
- Giegé, R., Ebel, J. P., Springer, M. & Grunberg-Manago, M. (1973) *FEBS Lett.* 37, 166-169.
- Goldstein, A. & Brown, S. J. (1961) *Biochem. Biophys. Acta* 53, 483-439.
- Grunberg-Manago, M. (1980) in "Ribosomes, Structure, Function and Genetics" (eds. G. Chambliss, G. R. Craven, J. Davis, K. Davis, L. Kahan & M. Nomura) University Park Press, Baltimore, pp. 445-477.
- Grunberg-Manago, M., Clark, B. F. C., Revel, M., Rudland, P. S. & Dondon, J. (1969) *J. Mol. Biol.* 40, 33-44.
- Harvey, R. J. (1973) *J. Bacteriol.* 114, 309-322.
- Hershey, J. W. B., Duncan, R., Etchison, D. O. & Milburn, S. (1984) in "Gene Expression. The Translational Step and its Control" (eds. B. F. C. Clark & H. U. Petersen), Munksgaard, Copenhagen, pp. 58-76.
- Hoagland, M. B., Zamecnik, P. C. & Stephenson, M. L. (1957) in "A Symposium on Molecular Biology" (ed. R. E. Zirkle), University of Chicago Press, pp. 105.

- Holley, R. W., Apgar, J., Everitt, G. A., Madison, J. T., Marquisse, M., Merrill, S. H., Penswick, J. R. & Zamir, A. (1965) *Science* 147, 1462.
- Jacques, Y. & Blanquet, S. (1977) *Eur. J. Biochem.* 79, 433-441.
- Jones, M. D., Petersen, T. E., Nielsen, K. M., Magnusson, S., Sottrup-Jensen, L., Gausing, K. & Clark, B. F. C. (1980) *Eur. J. Biochem.* 108, 507-526.
- Kahn, K., Fromant, M., Fayat, G., Dessen, P. & Blanquet, S. (1980) *Eur. J. Biochem.* 105, 489-497.
- Ladner, J. E., Jack, A., Robertus, J. D., Brown, R. S., Rhodes, D., Clark, B. F. C. & Klug, A. (1975) *Proc. Nat. Acad. Sci. USA* 72, 4414-4418.
- Léon, M., Dondon, J., Labousse, J., Grunberg-Manago, M. & Buckingham, R. H. (1979) *Eur. J. Biochem.* 98, 149-154.
- Lodish, H. F. (1976) *Amr. Rev. of Biochem.* 45, 39-72.
- Marcker, K. A. (1965) *J. Mol. Biol.* 14, 63-70.
- Marcker, K. A. & Sanger, F. (1964) *J. Mol. Biol.* 8, 835-840.
- Morikawa, K., la Cour, T. F. M., Nyborg, J., Rasmussen, K. M., Miller, D. L. & Clark, B. F. C. (1978) *J. Mol. Biol.* 125, 325-338.
- Morowitz, H. J. & Spaulding, M. (1958) *Biochem. Biophys. Acta* 29, 514-521.
- Nirenberg, M. W. & Matthaei, J. W. (1961) *Proc. Nat. Acad. Sci. USA* 47, 1588-1602.
- Ofengand, J. (1977) in "Molecular Mechanisms of Protein Biosynthesis" (eds. H. Weissbach & S. Pestka) Academic Press, Inc., New York, pp. 7-79.
- Ofengand, J., Gornicki, P., Nurse, K. & Boublik, M. (1984) in "Gene Expression. The Translational Step and its Control" (eds. B. F. C. Clark & H. U. Petersen), Munksgaard, Copenhagen, pp. 293-315.
- Petersen, H. U. (1980) Thesis of Dr. of Sciences, Université de Paris XI, The Academy of Paris.
- Petersen, H. U., Danchin, A. & Grunberg-Manago, M. (1976a) *Biochemistry* 15, 1357-1362.
- Petersen, H. U., Danchin, A. & Grunberg-Manago, M. (1976b) *Biochemistry* 15, 1362-1369.
- Petersen, H. U., Joseph, E., Ullmann, A. & Danchin, A. (1978) *J. Bacteriology* 135, 453-459.
- Petersen, H. U., Kruse, T. A., Worm-Leonhard, H., Siboska, G. E., Clark, B. F. C., Botorin, A. S., Remy, P., Ebcl, J. P., Dondon, J. & Grunberg-Manago, M. (1981) *FEBS Lett.* 128, 161-165.
- Petersen, H. U., Röhl, T., Grunberg-Manago, M. & Clark, B. F. C. (1979) *Biochem. Biophys. Res. Commun.* 91, 1068-1074.
- Petersen, H. U., Siboska, G. E., Clark, B. F. C., Buckingham, R. H., Hountondji, C. & Blanquet, S. (1984a) *Biochimie* 66, 625-630.
- Petersen, H. U., Wikman, F. P. & Clark, B. F. C. (1984b) in "Gene Expression. The Translational Step and its Control" (eds. B. F. C. Clark & H. U. Petersen), Munksgaard, Copenhagen, pp. 17-21.
- Petersen, H. U., Wikman, F. P., Siboska, G. E., Worm-Leonhard, H. & Clark, B. F. C. (1984c) in "Gene Expression. The Translational Step and its Control" (eds. B. F. C. Clark & H. U. Petersen), Munksgaard, Copenhagen, pp. 41-57.
- Pingoud, A., Urbanke, C., Peters, F. & Mass, G. (1977) *Eur. J. Biochem.* 78, 403-409.
- Plumbridge, J. A., Deville, F., Sacerdot, C., Petersen, H. U., Cenatiempo, Y., Cozzone, A., Grunberg-Manago, M. & Hershey, J. W. B. (1985) *EMBO J.* 4, 223-229.

- Rubin, J. R., Morikawa, K., Nyborg, J., la Cour, T. F. M., Clark, B. F. C. & Miller, D. L. (1981) *FEBS Lett.* 129, 177-179.
- Sacerdot, C., Dessen, P., Hershey, J. W. B., Plumbridge, J. A. & Grunberg-Manago, M. (1984) *Proc. Natl. Acad. Sci. USA* 81, 7787-7791.
- Sacerdot, C., Fayat, G., Dessen, P., Springer, M., Plumbridge, J. A., Grunberg-Manago, M. & Blanquet, S. (1982) *EMBO J.* 1, 311-315.
- Samuel, C. E., D'Ari, L. & Rabinowitz, J. C. (1970) *J. Biol. Chem.* 245, 5115-5121.
- Samuel, C. E., Murray, C. L. & Rabinowitz, J. C. (1972) *J. Biol. Chem.* 247, 6856-6865.
- Samuel, C. E. & Rabinowitz, J. C. (1974) *J. Biol. Chem.* 249, 1198-1206.
- Shine, J. & Dalgarno, L. (1975) *Nature* 245, 34-38.
- Sprinzl, M. & Gauss, D. H. (1984) *Nucl. Acids Res. Supp.* 12, r1-r57.
- Steitz, J. A. (1980) in "Ribosomes, Structure, Function and Genetics" (eds. G. Chambliss, B. R. Craven, J. Davis, K. Davis, L. Kahan & M. Nomura) University Park Press, Baltimore, pp. 479-495.
- Stöffler, G., Bald, R., Kastner, B., Lührmann, R., Stöffler-Meilicke, M. & Tischendorf, G. (1980) in "Ribosomes, Structure, Function and Genetics" (eds. G. Chambliss, B. R. Craven, J. Davis, K. Davis, L. Kahan & M. Nomura) University Park Press, Baltimore, pp. 171-205.
- Tanada, S., Kawakami, M., Nishio, K. & Takemura, S. (1982) *J. Biochem.* 91, 291-299.
- Waller, J.-P. (1963) *J. Mol. Biol.* 7, 483-496.
- Waller, J.-P. & Harris, J. I. (1961) *Proc. Nat. Acad. Sci. USA* 47, 18-23.
- Watson, J. D. (1964) *Bull. Soc. Chim. Biol.* 46, 1399-1425.
- White, B. N. & Bayley, S. T. (1972) *Biochem. Biophys. Acta* 272, 583-587.
- Wikman, F. P., Siboska, G. E., Petersen, H. U. & Clark, B. F. C. (1982) *EMBO J.* 1, 1095-1100.
- Wrede, P., Woo, N. H. & Rich, A. (1979) *Proc. Nat. Acad. Sci. USA* 76, 3289-3293.
- Zelwer, C., Risler, J. L. & Brunie, S. (1982) *J. Mol. Biol.* 155, 63-81.

JON STEEN PETERSEN

*The Directional Solidification of
Silicate melts: Crystallization Kinetics
and Macrosegregation*

ABSTRACT. Textures and compositional variations of minerals in the contact-zone of a plutonic magma body show that solidification occurs under continuously changing conditions. Changes in crystal morphology reflect successive stages of supercooling in the contact-zone, where initial thermal supercooling is followed by the sequential increase and decrease of constitutional supercooling. Unidirectional, columnar growth ceases with the eventual elimination of constitutional supercooling and is succeeded by the formation of cumulitic – equigranular rocks. Compositional variations in the solid products reflect competitive growth kinetics in a plane front solidification condition, and are similar to variations observed in many rhythmic layered rocks. As the thermal conditions favour heterogeneous nucleation and growth of crystals that are attached to the solidification front, a chain nucleation and growth mechanism is predicted that produces regular solid interfaces towards the liquid, capable of creating both rhythmic compositional variations and steady state solidification. Layering and other planar features thus become solidification isochrons. Differentiation of the solid product is related to the rate of advance of the solidification front. Rapid advance of the interface creates differentiated products because solute accumulation cannot be removed. A slowly advancing interface may allow effective adcumulus solidification, and thus preserve primitive compositions.

The analysis of dynamic solute redistribution during solidification bears analogy with processes of macrosegregation in metal castings. Notable features include the formation of negative and positive segregation in the lower and the upper portions of a single charge, respectively. The formation of regular compositional reversals, without relation to influx of new melt batches, emphasizes the potential of macrosegregation theory for the solution of important petrologic problems.

Department of Geology, University of Aarhus, Denmark

1. Introduction

The transformation of liquid magma into a solid product is a complicated process that involves rearrangement of liquid molecules into stable nuclei, transport of constituent elements to the nucleation site and removal of latent heat of crystallization from the growing phase. Each of

these processes may be a limiting factor for the solidification and affect the resulting product (Kirkpatrick, 1975). As the solid product is usually an aggregate of different mineral species, its overall composition will also depend on the rate of segregation of the individual phases during solidification. In order to solve the problems of magma solidification, therefore, knowledge of the rate limiting factor at any stage of the solidification is essential. Although information about stable phase assemblages allows prediction of possible differentiation paths, important restrictions exist because solidification conditions change continuously during cooling. The analysis of crystallization kinetics, which permit prediction of some effects of changing rate-controlling factors, is essential to the understanding of dynamic magma solidification processes.

Equilibrium phase relations are generally used to determine the stable solid phase assemblage of a given liquid, or the relative amounts of the constituent phases for a given temperature and pressure, (e.g. Morse, 1980). Natural rocks, however, often are not equilibrium assemblages because 1) their phase assemblages do not correspond to any reasonable source liquid (e.g. monomineralic rocks), 2) they represent disequilibrium assemblages (e.g. certain porphyritic rocks) or 3) their compositional variations are too rapid to be accounted for by changing bulk liquid compositions (e.g. rhythmic layered rocks; see review by Irvine, 1979). For all these cases, differential segregation of the constituent phases is inferred, and was traditionally considered to result mainly from physical processes of sorting such as gravity induced crystal settling (e.g. Hess, 1938; Wager & Brown, 1967) or fluid flow differentiation (Baragar, 1960; Bhattacharji & Smith, 1964; Komar, 1972). Recently, also kinetic factors have been emphasized as a prime source for fractionation during crystallization (McBirney & Noyes, 1979), the segregation being essentially chemical, caused by competitive nucleation and different growth rates in a static melt environment.

The kinetics relating to solidification of silicate melts have generated considerable interest in the latest decade (Lasaga & Kirkpatrick, 1981). Experimental studies on the crystallization of single phases have produced abundant data on nucleation and growth properties of individual minerals (Lofgren, 1974a, 1974b; Lofgren & Gooley, 1977; Donaldson, 1976; Fenn, 1977) whereas natural or synthetic rock melts have provided data on multiphase systems (Lofgren et al., 1974; Donaldson et al., 1975; Usselman et al., 1975; Walker et al., 1976; Lofgren, 1977, 1983; Grove & Walker, 1977; Swanson, 1977; Grove, 1978; Grove & Beaty, 1980; Schiffmann & Lofgren, 1982). These recent studies were preceded by

Harker (1909) who emphasized many aspects of crystallization kinetics, such as nucleation and growth control, influence of supercooling and faceted versus non-faceted growth. His contribution though remarkably perceptive, was overshadowed by the concept of equilibrium crystallization, and never gained the influence on petrological research it deserved.

A major problem in the application of many kinetic data to petrogenetic modeling of natural rocks is the complexity of the solidification geometry and the possible influence of fluid flow. In this respect, the study of directionally solidified rocks offers a unique opportunity to quantify some aspects of cooling and solute redistribution at a solid-liquid interface. In these rocks the product solidified sequentially along a lateral axis, thus, the diffusion problem is well defined.

Recognition of directionally solidified, natural rocks as a key to the understanding of magmatic processes was brought to prominence with the discovery and characterization of the Willow Lake layering (Poldervaart & Taubeneck, 1959; Taubeneck & Poldervaart, 1960). Later, similar features were discovered along the margins of minor intrusions in the US-Cordillera (Moore & Lockwood, 1973) and elsewhere. Perpendicular feldspars in the border zone of the Skaergaard intrusion, E. Greenland (Wager & Deer, 1939, Wager & Brown, 1967), are other examples of directional solidification, as are the spectacular harrisitic olivines and pyroxenes of the Rhum ultramafic complex (Brown, 1956; Wadsworth, 1960). Some dikes occasionally show elongated growth normal to the borders (Platten & Watterson, 1969; Drever & Johnston, 1972) as do many pegmatitic occurrences (Jahns & Tuttle, 1963). A link to the pegmatitic structures can be found in the remarkable layering and directional solidification textures recently described from granitic complexes associated with the Climax and Henderson Mo-ore bodies in Colorado (Shannon et al., 1982). Morphologically similar mineral textures have also been observed in rapidly chilled volcanic rocks such as the borders of pillow lavas (Bryan, 1972), in the spinifex-textured komatiites (Nesbitt, 1971; Donaldson, 1983) and in lunar materials (e.g. Donaldson et al., 1977).

A remarkable sequence of comb-textured rocks has been discovered recently along the contact of the lardalite nepheline-syenite complex, the Oslo Province, (Petersen, 1978a, 1978b, 1985). These form a 2-10m wide, marginal facies, with branching, dendritic minerals oriented normal to the contact and display compositional variations which include complementary phase-layering, within-layer compositional banding, columnar, monophasic precipitation and gradual transition from columnar, comb-textured rocks to laminar, porphyritic varieties. Beyond

these contact zones, the interior pluton consists of extremely coarse grained, cumulitic rocks with alkali feldspar – nepheline rich outer parts and increasingly olivine – pyroxene rich inner parts with a complete, gradational transition between these extremes (Table 2 in Petersen, 1978b). This compositional variation is reverse to that of the contact zones and clearly requires a different explanation than simple prograde fractionation in a decreasing temperature regime. The range of textures and compositional variations of the lardalite complex will provide the background for the present study of directional solidification and implications for compositional variations and macrosegregation of solidifying, granular or cumulus textured magmas.

2. Columnar solidification

2.1 General setting of comb-layered rocks

The contact zones of the Lardalite intrusion exhibit directional dendritic growth of olivine (Fo_{61}), Ca-pyroxene ($\text{En}_{40}\text{Fs}_{12}\text{Wo}_{48}$), ternary feldspars ($\text{An}_{16}\text{Ab}_{75}\text{Or}_8$ to $\text{An}_7\text{Ab}_{63}\text{Or}_{30}$), nepheline and Fe-Ti oxides. These minerals appear sequentially in the contact zones from the border towards the interior. Dendritic olivines and Fe-Ti oxides are enriched in the marginal facies with alkali feldspar, Ca-pyroxene and nepheline successively appearing inward. There is, however, a wide overlap in the appearance of the feldspars and pyroxene as indicated in Fig. 1.

The crystal shape varies systematically from the margin towards the interior; dendritic varieties of the major phases are gradually succeeded by faceted ones. The changes in phase assemblage and crystal shape clearly indicate that the conditions for formation of dendritic versus faceted crystal shapes are distinct for the different phases; some phases grow with euhedral shapes when others possess highly irregular dendritic shapes.

In addition to the perpendicular, dendritic crystal shapes, the contact zones possess conspicuous *compositional layerings* normal to growth which include changing phase assemblages as well as modal fluctuations within single layers. Well defined layers of dendritic feldspar or nepheline alternate with layers of curved, dendritic Ca-pyroxene, creating a type-1 layering (Petersen, 1985). Boundaries between successive layers of this type may be sharp or gradational (Fig. 2A). Another type of compositional banding is produced by periodic variations in the amount of interdendritic material along planes normal to growth direction. This

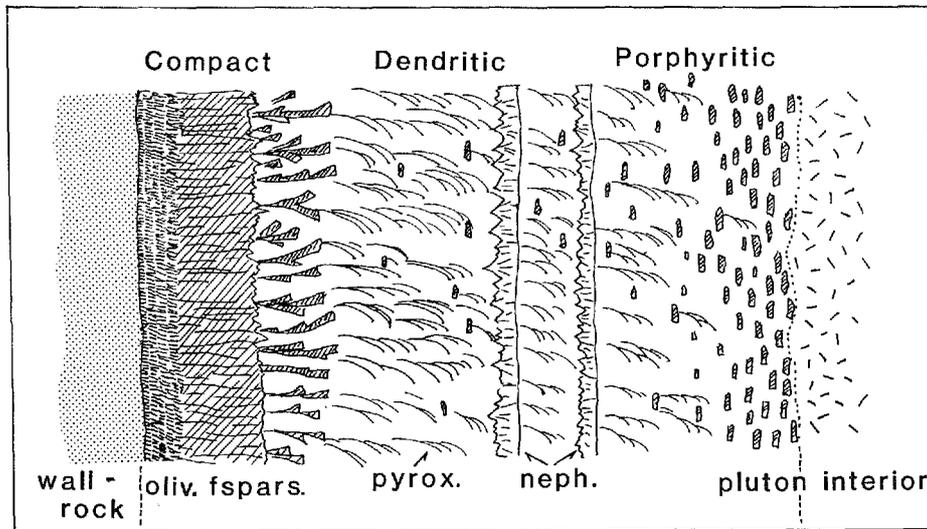


Fig. 1. Schematic cross section, showing major features of the lardalite contact zones. A marginal, olivine-rich dendrite zone is succeeded by compact and dendritic feldspar zones, which in turn are followed by dendritic pyroxene- and nepheline-rich zones. Porphyritic textures dominate the innermost portion towards the cumulus-textured pluton interior.

creates a compositional layering which typically fluctuates about constant composition and is termed type-2 layering (Fig. 2B).

Because of the large compositional variations, including some nearly monomineralic layers, average compositions of the contact zones can not be accurate estimates of the initial magma composition. Compositional variations of the individual minerals do, however, provide valuable information on the nature of the source liquid composition during solidification.

2.2 Crystal morphologies: feldspars

Ternary feldspars are the only minerals to have dendritic morphologies almost throughout the entire width of the contact zones, and also display the most diverse compositional and textural variations. For this reason, the ternary feldspars are the most useful for the study of the solidification conditions throughout the contact-zone formation.

The feldspars appear in four principal morphologies in the contact zones. Three of these are columnar with their longest crystallographic axis directed normal to the contact and the assumed solidification front, while the fourth consists of elongate, flattened and occasionally rhomb-

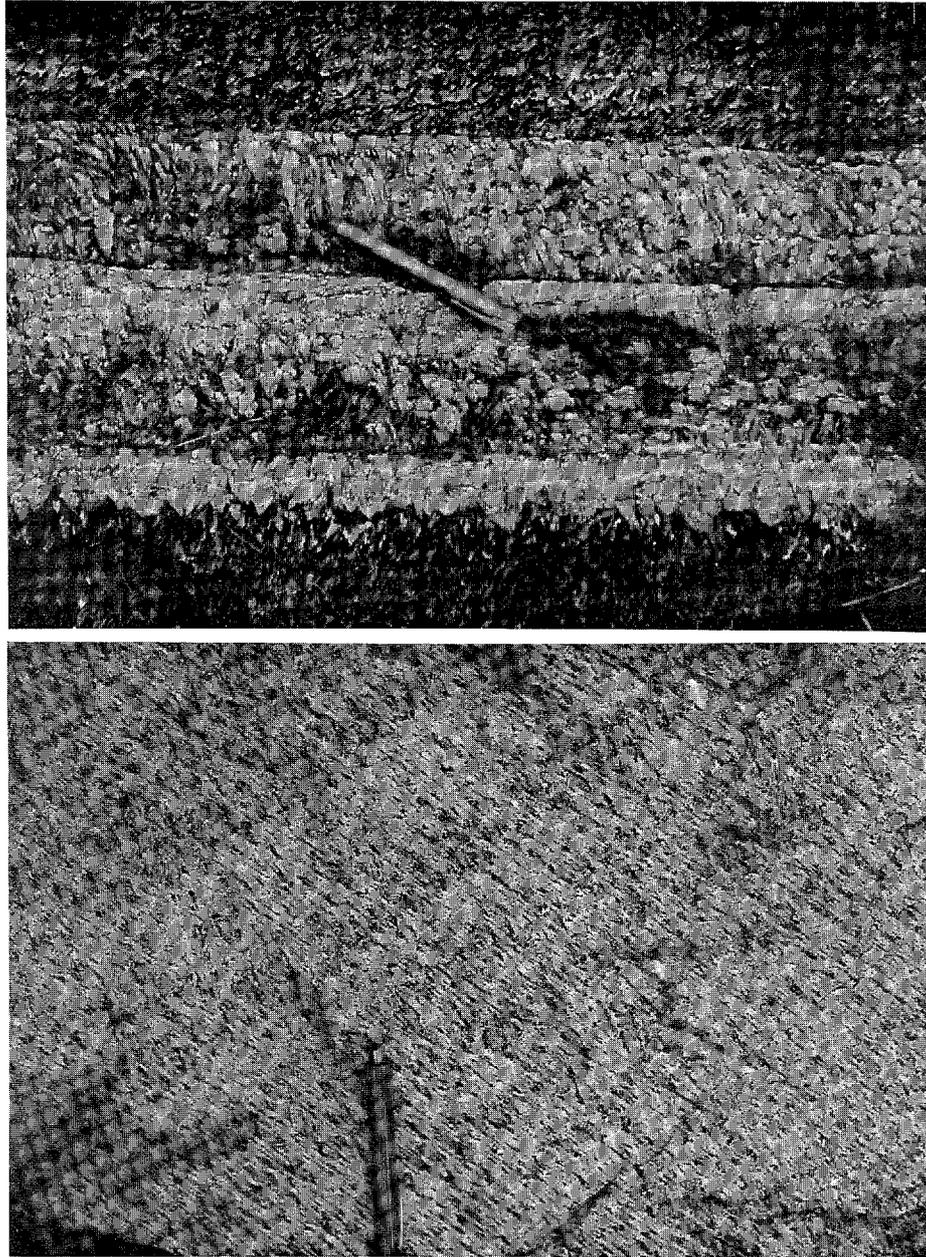


Fig. 2. A) Complementary phase layering. Dendritic pyroxene is succeeded by dendritic feldspars that gradually widen out to form compact layers. The upper boundaries of the feldspar layers are sharp whereas gradual transitions characterize the upper pyroxene boundaries. Growth is towards the upper part of the photo.

B) Compositional banding in columnar-dendritic feldspar layer. Compositional fluctuations mark small changes in steady-state solidification. Growth direction towards lower right corner.

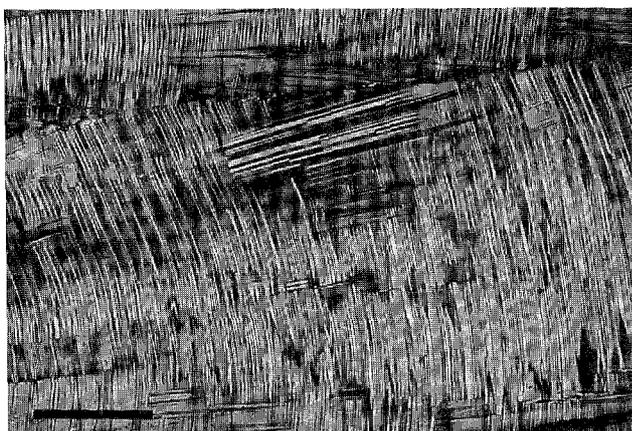
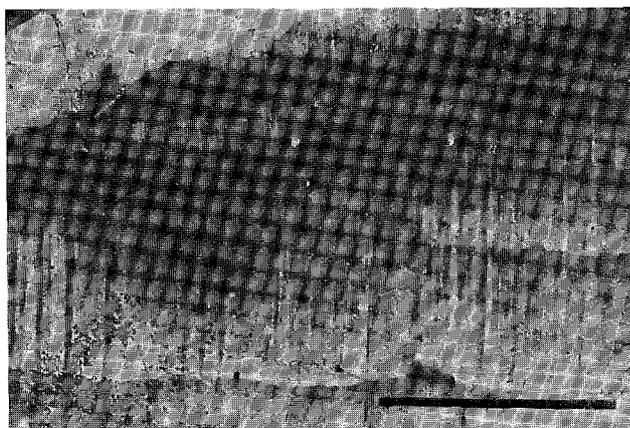
shaped phenocrysts with their longest axis parallel to the contact. The feldspar occurrences have been subdivided into the following zones: 1) picritic rim-zone; 2) monomineralic, columnar feldspar zones, 3) dendritic feldspar zones, and finally 4) porphyritic feldspar zones (Petersen, 1985).

Picritic rim-zone feldspars are low-angle branching or curving dendrites which expand in the growth direction. Universal-Stage observations indicate that the principal elongation is parallel to the optic z-axis and therefore normal to the 010 plane. The extinction of these feldspars is undulatory, sweeping across the subgrain upon rotation and parallel to the length of the crystal. Individual subgrains are separated by slightly off set extinction positions. Polysynthetic twinning is abundant, particularly in the transverse direction, and follows the albite-law (Fig. 3). Minor domains of pericline twins follow the longitudinal direction. The distinctly curved appearance of the transverse polysynthetic twins indicates that the undulous extinction across individual dendrite branches is related to curved lattice planes, which are disrupted at neighbouring subgrain boundaries. The slightly irregular, fan-shaped dendrites of the rim zone typically measure 10-20mm and their compositions cluster around $An_{16}Ab_{75}Or_9$.

Monomineralic feldspar zones occur at different levels in the contact zones. In one such zone, near the external contact of the pluton, columnar textured feldspar has developed into a more than 2 m wide, compact layer. The feldspars within this zone are elongated perpendicular to the contact and constitute a cellular substructure. Within this substructure numerous, minor subgrains possess almost identical orientations. Optic z-directions are parallel to the principal elongation; the maximum growth direction is therefore normal to the 010 plane. Thin sections cut parallel to the columnar fabric show that the subgrain boundaries form arrays of extinction discontinuities, which divide the aggregate into parallel lamellae. Extinction sweeps simultaneously from one side of the subgrain to the other (Fig. 4A). This shingle-like texture is identical to lineage structures which are well known from unilaterally solidified metals (Buerger, 1934), and are presumably caused by accumulated dislocations along subgrain boundaries during rapid, unilateral solidification (Chalmers, 1964). Within individual subgrains of this class of feldspar, the curved nature of the lattice is revealed by a systematic change in the orientation of the optic axes across the width of a single grain, typically by 5-10° but up to 20-25° (Petersen, 1985). This is equivalent to that exposed by the curved, transverse Ab-twins in the picrite-zone feldspars.

Fig. 3. A) Transverse, polysynthetic albite-twins in marginal picrite-zone feldspars. Twin plane curvature reflects curved lattice planes of the host feldspar. Growth direction towards right. Scale bars are 0.5 mm.

B) Synthetic reproduction of the curved, polysynthetic twins in feldspar-analogue material (Suberic acid). The crystals were grown directionally in high supercooling before quenching and formation of martensitic twins. (Both albite and pericline law twins).



Thin sections cut normal to the main growth direction reveal a cellular substructure with rhomb-shaped subgrain patterns of 0.7×1 mm and a spectacular, plaited, mesoscopic, mosaic structure (Fig. 5). Nearly identical orientation of the individual subgrains makes cleavage surfaces appear continuous like a huge single crystal and only the slightly curved appearance of each subgrain makes it distinguishable from its neighbour. The plaited mosaic structure represents two symmetrically arranged orientations divided by a coherent interface. This interface formed during primary growth by the continuous adjustment of neighbouring cell colonies into a minimum-energy boundary configuration such as twin planes as shown by Universal-Stage studies (Petersen, 1985). The rhomb-shaped pattern is a result of the anisotropic lattice properties of the monoclinic feldspars compared to the honeycomb-shaped cell

boundaries observed in many directionally solidified metallic compounds which possess isotropic, non-faceted surface properties (Buerger, 1934; Rutter & Chalmers, 1953).

The inner two-thirds of the border zones are comprised of several varieties of dendritic feldspar morphologies. With the exception of a few compact, cellular feldspars in the monomineralic zones, the dendrites in the inner parts are generally columnar with a finite spacing. The interdendritic material is usually fine grained, equigranular aggregates of complementary phases to the dendritic mineral.

Two principal dendrite types can be distinguished in this part of the contact zone. One type consist of columnar, non-branching feldspars, with constant or nearly constant spacing and orientation perpendicular to the contact or previous phase-layer contact (Fig. 6). The other, gener-

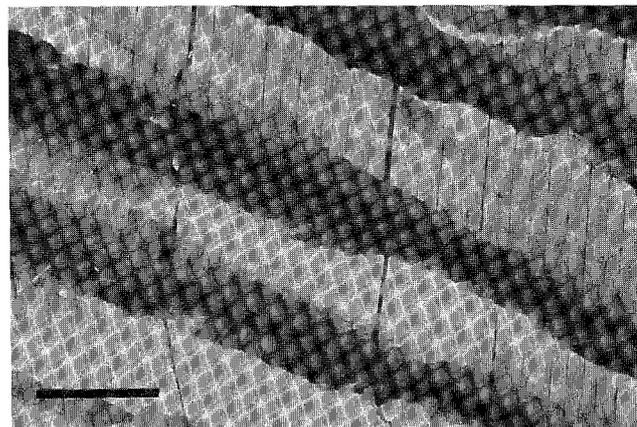
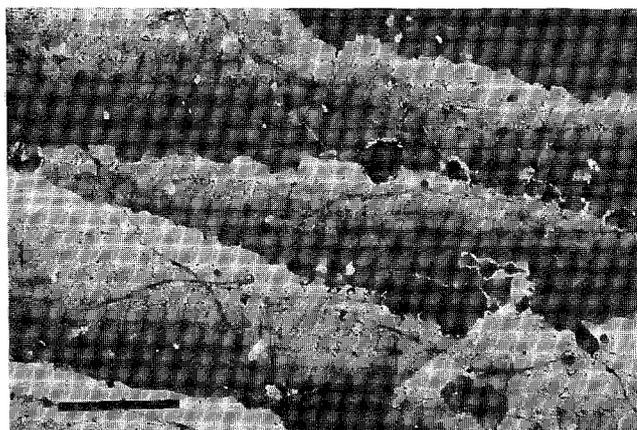


Fig. 4. A) Lineage textures in the compact feldspar zone. The shingle-like texture is due to the undulose extinction of curved lattice planes and arrays of grain boundary misfits. Growth direction towards the left. Bar 0.5 mm. B) Synthetic shingle texture in feldspar analogue materials, produced at moderate cooling rate. The structure is formed during growth by the advancement and reproduction of low-angle lattice misfits (lineage structures).

Fig. 5. Plaited structure in compact feldspars viewed normal to the main growth direction. The structure reveals a regular arrangement of subgrains or cell colonies along two dominant directions, controlled by the crystallographic properties of the growing phase.

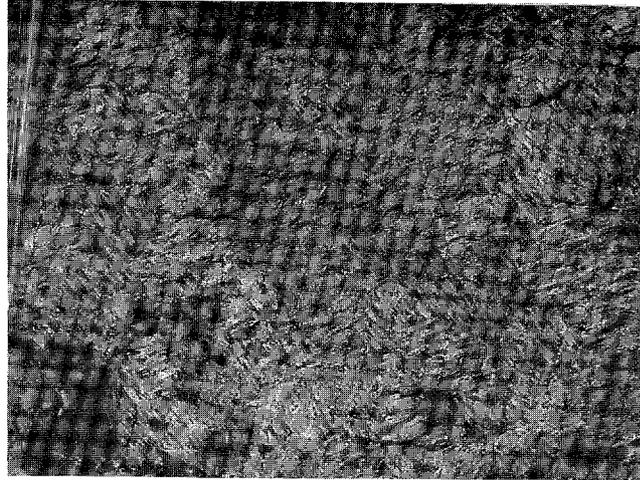


Fig. 6. Columnar-dendritic feldspars with constant spacing, indicating steady-state, diffusion controlled growth.

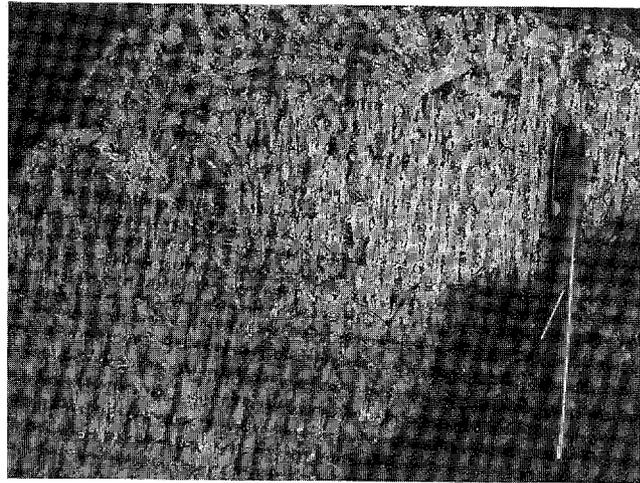


Fig. 7. Faceted feldspar dendrites have their maximum elongation deflected from the perpendicular direction to the contact. Second-arm branching compensates for the wide dendrite spacing and show maximum elongation parallel to the isotherms.

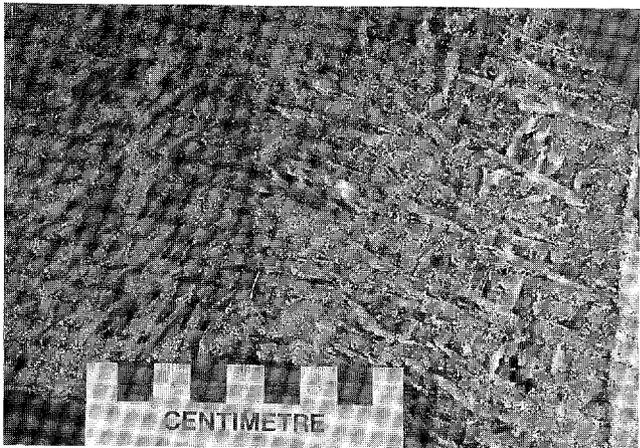




Fig. 8. Laminar textures from the inner contact zones are produced by the parallel orientation of euhedral feldspar phenocrysts with their maximum elongation parallel to the isotherms and towards the viewer.

ally occurring further inwards, is more widely spaced and oriented with the maximum elongation at a finite angle to the contact-normal direction (Fig. 7). Because of the wide spacing of these non-orthogonal dendrites, second arm branching is occasionally developed with a maximum elongation along the cooling front instead of normal to it. This texture marks a transition to the innermost zone, where feldspar phenocrysts with lamellar textures are oriented parallel to the cooling front.

The innermost portions of the contact zones are distinctly porphyritic and consist of euhedral, rhombshaped feldspar phenocrysts of 2–4 cm length set in a fine- to medium-grained matrix of olivine, pyroxene, nepheline, and biotite. The proportions of the matrix phases may vary highly resulting in ultramafic to foyaitic matrix compositions. The feldspars show consistent, parallel orientations with their longest axis parallel to the contacts and imprint a pronounced laminar textural fabric to the rock (Fig. 8).

Dendritic pyroxene and nepheline often nucleate and grow on the larger feldspar facets, indicating that substantial constitutional supercooling gradients exist in the matrix liquid. This in turn, suggests that deposition of the feldspars was not a result of gravitative settling or fluid flow along the solidification front but rather precipitation in a static boundary layer where conditions for constitutional supercooling persisted only for pyroxene and nepheline; feldspar crystallization was near liquidus conditions. Supercooling is equivalent to supersaturation in this respect; therefore the contrasting mineralogy can be caused by supersaturation rather than supercooling in the boundary layer.

Olivines

Dendritic olivines occur exclusively in the outermost portion of the contact zones, the picritic rim-zone. These olivine dendrites are fairly short (2×6 mm) and display irregular bifurcation, mainly as low-angle, non-crystallographic branching in the preferred growth direction. Complex intergrowths occur primarily with Fe-Ti oxides. The co-existing feldspars tend to form isolated grains, with strongly embayed boundaries towards the olivine-oxide aggregates.

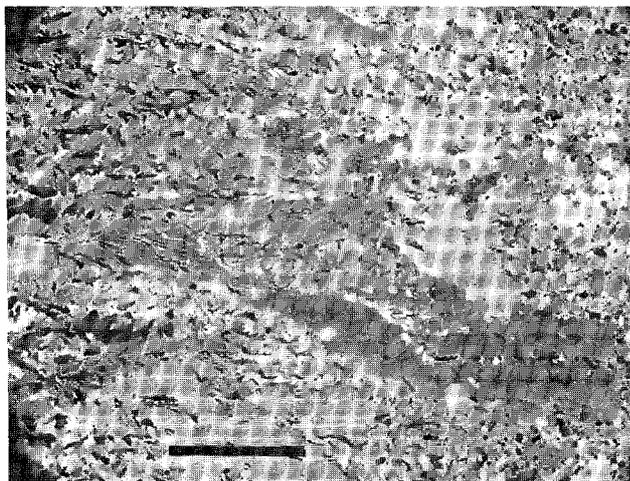
The textural relations show that the olivines transgress and overgrow the Fe-Ti oxides, while the feldspars in turn partly transgress the olivines (Fig. 9). Because of the unidirectional solidification texture, the clear separation of feldspar and olivine-oxide dendrites indicates that the former cannot have formed during subsequent, interstitial growth, but must have grown simultaneously with plagioclase feldspar. Plagioclase dendrites possibly had a higher growth-rate and thus, could engulf growing olivine dendrites.

In the remaining parts of the contact zones, olivine occurs exclusively as minor, stubby crystals, generally as part of the fine grained, interdendritic matrix and often as clustered or glomerophyric aggregates.

Pyroxenes

Pyroxenes are absent in the marginal olivine-dendritic picrite zone, but are abundant in the other portions of the contact zone. Almost exclusively the pyroxenes form curved, open-dendritic crystals of 2 to 150 mm length, invariably trending inwards-downwards towards the center of

Fig. 9. Feldspar - olivine - Fe-Ti-oxide dendrites from the picrite zone show complex intergrowth textures. The textural relations suggest that the oxides precede olivine and feldspar dendrites, which apparently grew simultaneously, side by side. Dark gray phase is plagioclase; white is olivine. Bar measures 1 mm.



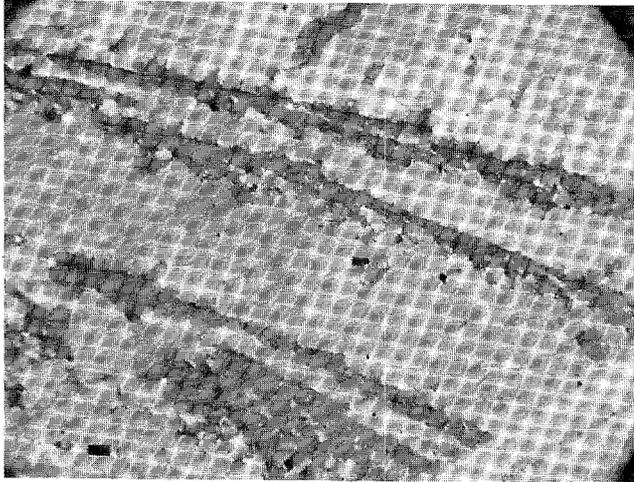


Fig. 10. Curved dendritic pyroxenes with low-angle, non-crystallographic branching along growth direction. The scale measures 1 mm.

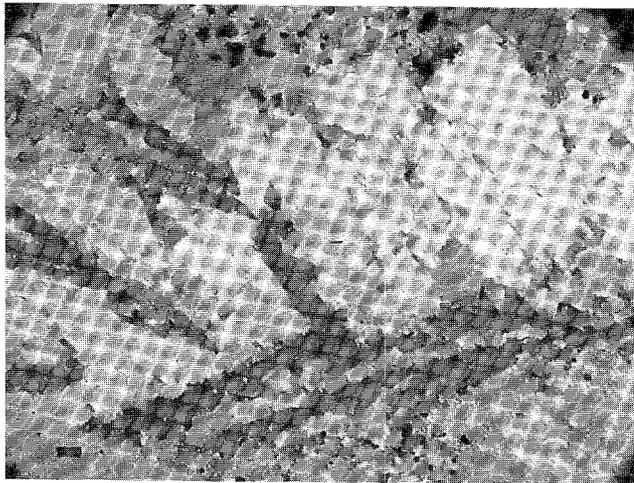
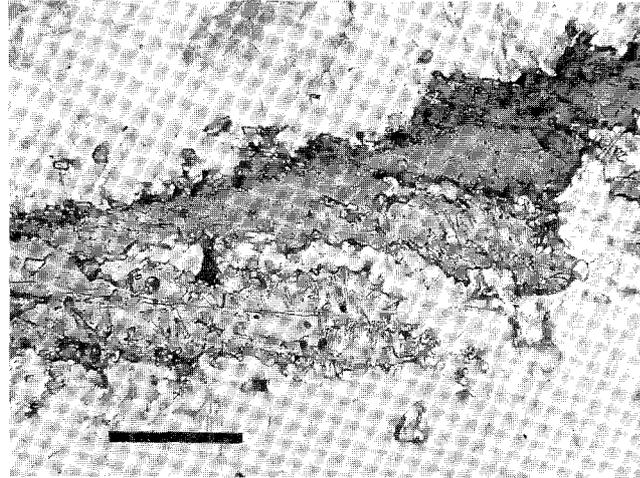


Fig. 11. Secondary nucleation of pyroxene on dendrites whose main elongation is at a low angle to the heat flow. White minerals are compact-dendritic nephelines with their growth directed towards the maximum heat flow. Bar-scale 1 mm.

the pluton (Fig. 14). Two broad types of bifurcation are present. One type shows continuous low-angle branching along the growth direction, that is partly related to the curved growth as it develops only on the concave side of the crystal (Fig. 10). The other type of bifurcation appears when the curved crystal reaches a maximum deviation from the perpendicular direction to the cooling front. In this case, subsequent branching occurs on the convex side of the crystal, initially perpendicular to the cooling front, but eventually curving (Fig. 11). This creates a chain of equally curved pyroxenes growing at large towards the cooling front (Fig. 13). These chains may actually constitute a single, continuous

Fig. 12. Minor apatite needles, included in the pyroxene dendrites, show consistent, parallel orientations. The apatite crystals have their maximum elongation parallel to the isotherms like the laminar pyroxenes and feldspars in Fig. 16. Bar is 1 mm.



crystal for several meters, although the trace of an individual stem is difficult to verify. The curvature may be steep, up to 80° of a circle with a radius of 3-4 cm, or be very gentle.

The pyroxenes form exclusively open dendrites with appreciable amounts of interdendritic materials, mainly consisting of complimentary phases to the pyroxenes such as feldspar, nepheline, olivine, and biotite. In sections normal to the maximum growth direction, some degree of lateral spreading occurs and the curved-dendritic structure may grow into a horizontally elongated, fan-shaped sheet. This sheet has a slightly convex upper surface and a ragged multibranched lower boundary.

Euhedral pyroxenes occur only in the innermost portion of the contact zones and often together with euhedral feldspar phenocrysts. These pyroxenes are oriented with their longest axis parallel to the cooling front. Nepheline is the only dendritic phase in this part and exhibits curved branching and compact, fan-spherulitic morphologies (Fig. 16).

Abundant minute, needle-shaped apatite crystals are included in most dendritic pyroxenes. The regular orientation of these apatite crystals suggests that they were continuously precipitating along a faceted interface of the growing pyroxene and constitute an important clue to the growth mechanism of the pyroxenes. Apatites included in the interior pyroxene stems are oriented parallel with their long axis normal to the maximum growth direction of the pyroxene (Fig. 12), whereas along the margin of the pyroxenes apatite needles are aligned parallel to the surface, suggesting a later formation when facets were formed also during

lateral growth. These apatites formed as a result of local supersaturation of P_2O_5 in the liquid boundary-layer ahead of the growing pyroxenes, where residual components were accumulated. Surprisingly, similar amounts of apatite crystals are never encountered in feldspar dendrites formed simultaneously with the pyroxene dendrites. It indicates that important differences exist in the diffusive boundary layer adjacent to the different minerals.

Nepheline

Nepheline occurs in the innermost two-thirds of the contact zones, and gradually increase in abundance towards the center. The primary morphology of this phase is compact spherulitic with multiple low-angle branching and widening in growth direction, generally normal to the contact and overall cooling front.

Compact spherulitic and columnar dendritic nephelines characteristically nucleate on the upper surface of earlier dendrites or phenocrysts and in particular, on the convex side of curved-dendritic pyroxenes. These relationships imply a relatively late formation. In the inner feldspar-porphyrific and laminar-textured zones abundant growth of dendritic nepheline towards the magma interior, i.e. normal to the contact and the lamination fabric, indicate a prominent supercooling/supersaturation gradient. No textural evidence of this gradient is recorded by the feldspar morphology (Fig. 15).

Nepheline produces compact monomineralic layers, due to the trans-

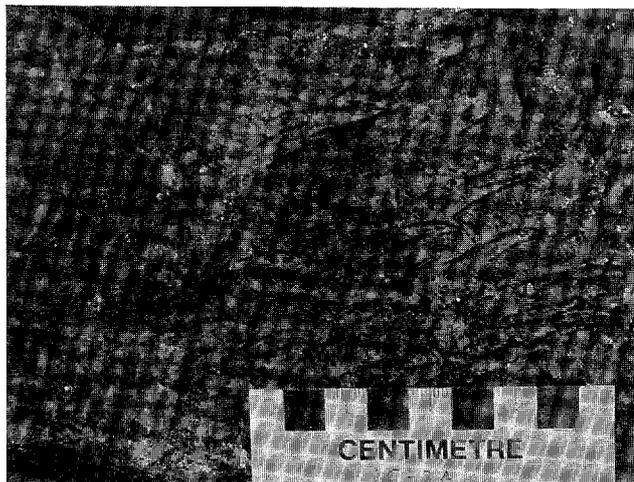


Fig. 13. Dendritic pyroxenes with repeated nucleation along the growth direction.

Fig. 14. Ca-pyroxenes with typical, curved dendritic appearance. White crystals are nephelines growing directionally on the upper part of the pyroxenes (see Fig. 11 for petrographic appearance).

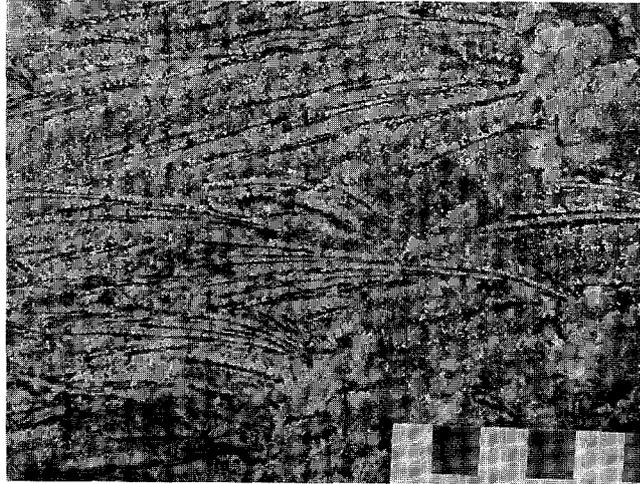


Fig. 15. Dendritic pyroxenes and nephelines in a feldspar porphyric zone. Nepheline is white, feldspar gray and pyroxene black. The feldspar phenocrysts measure 1-2 cm.

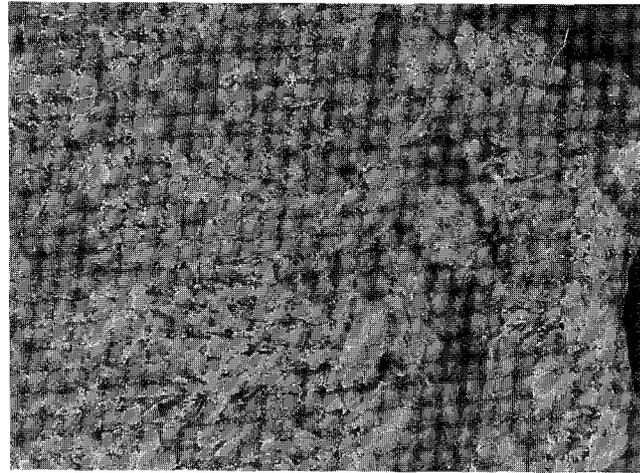


Fig. 16. Curved nepheline dendrites in a zone where pyroxenes and feldspars form faceted phenocrysts.



gressive nature of its growth. Transitional zones where cellular nepheline increases gradually at the expense of typically open-dendritic pyroxene ultimately leading into compact Ne-layers, 2–40 cm wide, are occasionally found. A contact-perpendicular cellular fabric, similar to the compact feldspar zones previously described, is formed in these layers. Unlike the feldspar zones, very few oxide or other inclusions are recorded in the dendritic nepheline.

There is an interesting periodic nature in some of these compact nepheline layers. In one example three consecutive layers of compact nepheline, interlayered by dendritic pyroxene, systematically increase in width along the growth direction. The width of the nepheline layers increases from 1.5 cm to 4 and further to 11 cm while the intervening, open dendritic pyroxene zones measure 20 cm, 40 cm and 64 cm respectively. The intervening zones are remarkably homogeneous in composition and contain about 10% vol. pyroxene. This particular three-layer zone can be traced for at least 3 km along the contact which shows its strongly pervasive nature.

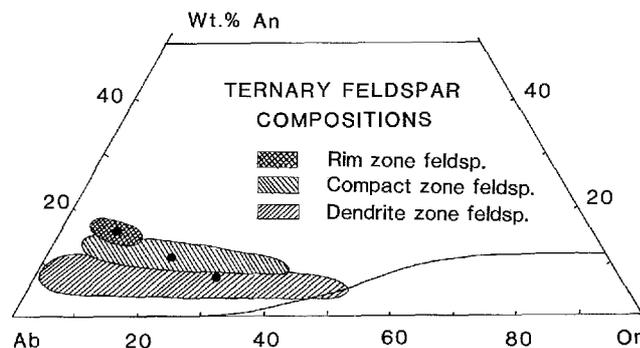
2.3 Mineral compositions

Chemical compositions of the dendritic minerals were determined by electron microprobe analysis using energy-dispersive methods (Petersen, 1985). Preliminary results are summarized in Figs. 17 and 18, which show the analyses of felsic and mafic phases respectively in the three major dendrite zones.

Exsolution into Na- and K-rich components forms a very fine, string-antiperthite, potash feldspar being subordinate to albite. A coarse segregation into a K-rich and a Na-rich feldspar is apparent along the margins of mesoscopic subgrain cells (Fig. 19a). This segregation is possibly a primary growth feature and the result of cellular solidification in which the subordinate, K-rich feldspar co-precipitated as an intercellular matrix, encouraged by the rejection of solute at the cellular front. Lofgren & Gooley (1977) experimentally produced similar intergrowth textures by growth from the melt (Fig. 19b). The process is somewhat analogous to the growth of non-coupled eutectic products (Elliott, 1977). An origin by structurally controlled, solid-state segregation is, however, also possible (Petersen, 1985).

Na and K content of the feldspar increase and the Ca content decreases, with increasing distance from the contact of the pluton. Compositions of rim-zone feldspars cluster about a common value while feldspars from the inner zones show increasingly Ca-poor compositions

Fig. 17. Composition of dendritic feldspars from the lardalite contact zones; cross-hatched field is picrite-zone feldspars; diagonally ruled fields are compact and dendritic feldspar zones, showing decreasing Ca-content towards the pluton interior. Solid dots mark the integrated composition of each type. The spread in Na- and K-rich components can be a primary growth feature.



and a broader spread in alkali-content. The compositions of the inner dendrite feldspars suggest the existence of Na- and K-rich end members in agreement with observed antiperthitic textures.

The primary feldspar composition of each group is given by the mean value of all analyses within a group (Fig. 17). The compositional variation towards the center of the pluton is consistent with a gradually decreasing temperature. No systematic variation in Ca has been detected within individual zones. The liquidus isotherms for 5 Kbar pressure (Morse, 1970), suggest that at this pressure, solidification temperatures would be in the 900° to 780°C-interval. Actual pressures during crystallization were probably slightly lower, about 3 kbar, thus these temperatures are conservative estimates.

The compositions of dendritic pyroxenes show limited variation and consistently plot around $En_{40}Fs_{12}Wo_{48}$. The Ca-rich nature of these pyroxenes is attributable to the initial alkali-rich melt composition and the lack of substantial amounts of co-precipitating plagioclase (Morse, 1979). These compositions (Fig. 18) are significantly more enriched in Al, Ti and Mg than typical larvikite and lardalite pyroxenes (Neumann, 1976). The dendritic pyroxene compositions from the different portions of the contact zones, however, are remarkably uniform.

Dendritic olivines are Fo_{62} whereas the granular varieties of the remaining contact zones are increasingly Fe-rich to about Fo_{45} . For comparison, the range of composition in typical larvikite and lardalite olivines is Fo_{45-30} and Fo_{55-30} respectively (Fig. 18). The more Mg-rich compositions of the olivines of the contact zones suggest a higher temperature origin than those of the typical lardalite.

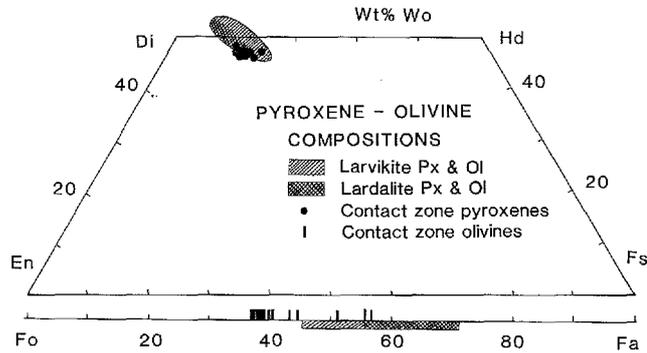
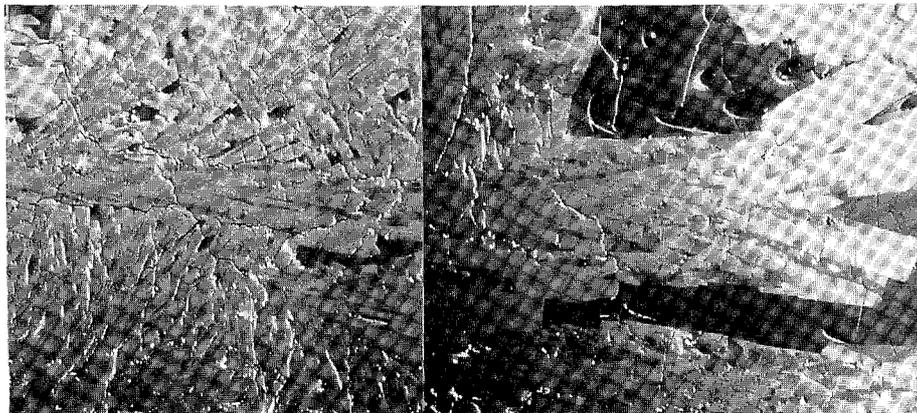


Fig. 18. Composition of dendritic pyroxenes and olivines in the lardalite contact zones, compared with average larvikite (diagonal ruling) and lardalite (cross hatched) pyroxenes and olivines. The contact zone minerals are more primitive than granular and cumulus textured minerals of the pluton interior.

Fig. 19. A) Feldspar dendrite showing segregation into Na-rich core (light) and K-rich margin (dark). Crossed nicolls; bar measures 2 mm.

B) Composite feldspar dendrite grown from synthetic ternary feldspar melt (Lofgren, pers. comm.). Separation into a K-rich (light) and Na-rich part (dark gray) is the result of growth from the melt. (backscattered electron image; total length of crystal: 6 mm.) Striking similarity with the structure of the feldspar dendrite in A) makes a related origin likely.



3 Discussion

3.1 Crystal morphologies

The relationship between growth morphology and solidification rate for silicate minerals has been intensively studied by experimental methods in the recent decade (Lofgren, 1974a; 1980; Kirkpatrick, 1981). These studies show that crystal morphology is strongly related to degree of supercooling/supersaturation; faceted solidification occurs with small deviations from equilibrium, whereas with increasing driving force, non-faceted and dendritic growth becomes important (Kirkpatrick, 1975, 1981). The main condition for dendrite formation is the presence of a supercooled liquid ahead of the solidification front, which makes planar interfaces unstable (Rutter & Chalmers, 1953; Jackson, 1958; Lofgren, 1974a). This condition develops when heat is extracted from the liquid faster than the rate of solidification (thermal supercooling), or when the liquidus depression of the melt attains a steeper curvature than the thermal profile towards the solid-liquid interface, because of progressive solute enrichment (constitutional supercooling) (Tiller et al., 1953).

The directional solidification textures of contact zones such as the lardalite intrusion, belong to a special class of igneous textures which have been termed thermotactic (Rinne, 1926; Drever & Johnston, 1972), crescumulate (Wagner et al., 1960), Willow Lake layering (Poldervaart & Taubeneck, 1959) and comb-textures (Moore & Lockwood, 1973). Comb-texture has been adopted here because of its ungenetic connotation and purely descriptive character.

Comb-textures are the result of crystal growth from the melt (Lofgren & Donaldson, 1975). Experimental studies of feldspar and pyroxene crystallization show that directional dendritic growth with branching and bifurcating morphologies, forms readily at moderate to high degree of supercooling (Lofgren, 1980). The curved appearance of some dendrites has been suggested to result from action of fluid flow near the crystallization front (Moore & Lockwood, 1973). However, curved and spiral growth occur in environments where fluid flow is nonexistent such as the two-dimensional growth of ice-dendrites, and thus most likely are the result of purely kinetic factors.

A main characteristic of the comb-texture is nucleation on a planar crystal interface, usually along the solid wall-rock of the intrusion, but also on freely suspended crystals. Heterogeneous nucleation on a foreign substrate requires only a fraction of the constituent atoms that is necessary for a critical size nucleus to form within the melt and therefore

occurs much more readily than any other nucleation event (Shewmon, 1969). During subsequent growth after nucleation, only crystals which are oriented with their fastest growing direction along the supercooling gradient will survive the initial growth transient. Crystals in other orientations will lag behind the growth front and ultimately become extinct (Donaldson, 1977). Repeated formation of low-angle subgrain branching during progressive solidification allows continuous lattice adjustment of the growing minerals to the optimum orientation. The process of reorientation on initial crystals therefore need not involve any subsequent nucleation.

The major growth mechanisms are continuous growth and growth by layer spreading (Dowty, 1980). During continuous growth all parts of a nucleus surface constitute potential sites of atomic attachment; growth is independent of crystallographic directions. Growth by layer spreading is related to the formation of crystal facets through lateral spreading in a crystallographic plane (Kirkpatrick, 1975). This mechanism has a lower increment rate in being restricted to a certain plane, and thus it has a slower growth rate than that of continuous growth.

The formation of facets is related to the crystallographic properties of the solid and the nature of the solid-liquid transformation. Growth morphologies at small degrees of undercoolings have been successfully predicted by the application of the so-called α -factor (Jackson, 1958):

$$\alpha = M \times (\Delta H_m/kT_e) \text{ or } = M \times (\Delta S_m/k)$$

where ΔH_m is heat of fusion change; k the Boltzmann constant; T_e the melting temperature; ΔS_m entropy of fusion; and $M = \eta/v$, the fraction of binding in the growth plane to total binding energy, also called the anisotropy factor. The α -factor thus consists of two terms, one being related to the entropy of fusion for the particular substance (ΔS_m) and the other being related to crystallographic properties of the substance (M). This shows an important property of facet formation, since it may vary both with respect to substance and with respect to crystallographic directions.

Most metals possess low values of ΔS_m which make their α -factor less than 2, implying growth by non-faceted, rough interfaces, i.e. a continuous mechanism (Jackson, 1972). Many organic materials have values greater than 2 and thus produce faceted interfaces upon freezing (Jackson & Hunt, 1966). The α -factors of most silicates are generally well above 2, except for quartz, albite, and sanidine which have low ΔS_m -values of

0.33, 0.75 & 0.77, respectively (Carmichael et al., 1974), and therefore α -factors about 2.

A interesting example of faceted growth is the crystallization of water, because ice has an α -factor near the critical value of 2 (~ 1.97) (Jackson, 1958). Because slightly higher binding energies are associated with the basal plane, this plane possesses α -values greater than 2, whereas other crystallographic planes have lower values. At small undercoolings, where growth anisotropy is maximized, facets form only along the basal plane; dendritic growth is experienced in all other directions. Freezing under these circumstances creates large flat sheets with maximum extension along the basal plane, (i.e. dendritic growth constrained by only one facet plane). Lateral growth of ice dendrites produces a cellular texture consisting of parallel sheets (Rohatgi & Adams, 1967) that completely mimics the braided texture of the compact-zone feldspars from the lardalite intrusion (see Fig. 5).

Under conditions of extreme supercooling, feldspar and pyroxene tend to form spherulitic masses, constituting aggregates of numerous crystals which radiate away from a common center (Keith & Padden, 1963). Technically, spherulites are multicrystalline aggregates, since each needle has an independent orientation; physically, however, all needles derive from a common nucleation center and thus constitute a single unit. The radiating nature is due to the presence of a constitutionally supercooled liquid around the freely suspended nucleus and the formation of abundant low-angle non-crystallographic branching of subgrains during subsequent constrained growth (Keith & Padden, 1963).

Two major categories of spherulites are defined (Keith & Padden, 1963) as compact (massive) and open (spiky) spherulites. In compact spherulites, neighbouring subgrains touch and leave no interdendritic material whereas the open spherulites contain appreciable amounts of interdendritic material. Miller (1977) found that in highly purified, monophase systems, the formation of facets was critical for the formation of open spherulites, as opposed to compact ones. A faceting plane will, because of growth competition, ultimately become oriented parallel to the maximum growth direction, and this facet will then allow separation from the neighbouring subgrain of the spherulite (i.e. formation of an open structure).

The morphologies of feldspar and nepheline in the monomineralic zones are analogous to that of compact spherulites, except that growth is not radiating from a common nucleation center but from a common nucleation plane in a parallel fashion. Initial radiating fabrics are occa-

sionally seen but are obliterated as adjacent spherulite colonies interconnect to produce a compact, parallel-textured layer (Fig. 2A). Because of the cooling geometry with isotherms parallel to the contact and heat of crystallization being extracted through the dendrite stem, a zone of constitutional supercooling will constrain the maximum growth normal to the interface (Rutter & Chalmers, 1953).

Feldspars

The consistent orientation of feldspars in the contact zones show that maximum elongation is invariably normal to the 010 plane (Petersen, 1985). This suggests that solidification is not a continuous mechanism, as usually assumed for cellular growth, but faceted with the orientation being controlled by at least one crystallographic direction. The lack of inclusions in the central part of the feldspar dendrites indicates that maximum growth occurred in the central part of a crystal face and spread towards the edges, possibly by a spiral growth mechanism caused by screw-dislocations. Other mechanisms involve nucleation of new layers along the edges and the formation of skeletal morphologies and inclusions in the central portion of the facets (Kirkpatrick, 1981).

Gilmer (1977) presented experimental evidence which supports computer simulations of crystal growth in showing that closely packed faces are more likely to form facets because of the higher binding energy associated with such planes. More loosely packed directions on the other hand maintain a rough interface at the lowest supercooling (i.e. they have high roughening temperature). The roughening temperature indicates the supercooling necessary for a particular crystallographic direction to transform from a slow growing facet to a rough, continuously growing interface (Gilmer, 1977; Hartmann, 1982).

In feldspar, the closest packed facet-plane is the 111 plane following Voensdrecht's (1983) analysis of PBC-vectors for plagioclase feldspar, using the crystal-bond concept of Hartmann-Perdock (Hartmann, 1973). The 010 plane is the most loosely packed f-face. According to Gilmer's (1977) results, but contrary to the common view, facets are likely to form initially on the most densely packed faces like 111 and proceed sequentially to the least densely packed faces like 010 in a sequence of decreasing supercooling and following the increasing roughening temperatures of various facets. This mechanism will allow the 010 plane at some stage during solidification to be the only face to maintain a high growth rate because of its relatively high atomic roughness. The com-

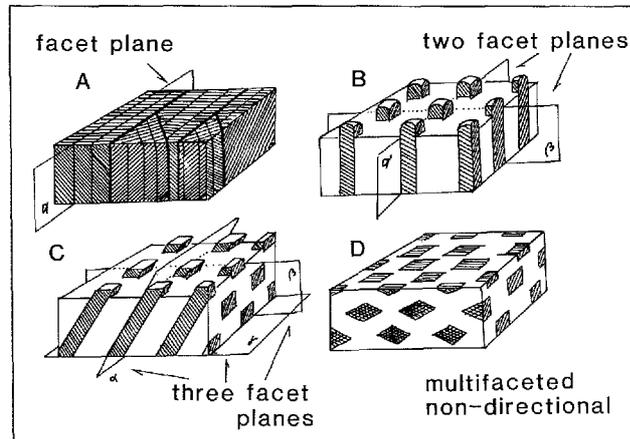
petitive nature of directional solidification will therefore allow only this direction to grow normal to the isotherms.

By analogy with the ice discs discussed above, the growth of a single facet, once formed, will be retarded, and ultimately become oriented normal to the fast growing direction due to the continuous competition for maximum growth rate. A single facet will restrict crystal growth in only one direction, resulting in parallel sheets with similar orientations in compact spherulites (Fig. 20A). Initially, individual sheets may show random orientations normal to the faster growth direction, but because of competitive growth, they ultimately become aligned to form a superficial sheet structure. Twin planes may allow modification of this sheet structure to form a braided texture of two or more directions as observed in the compact, columnar feldspar zones (Fig. 5).

When two facets form, growth is restricted to two directions in space. This allows formation of columnar structures which are separated by varying amounts of interdendritic materials in the growth plane (analogous to rod eutectics) (Fig. 20B). Maximum growth is largely unconstrained in the third direction which is still below its roughening temperature and accordingly oriented parallel to heat flow, i.e. perpendicular to the contact with the magma. At this stage the compact feldspar structure will grade into a columnar texture with interdendritic material becoming abundant towards the inner portion of the contact zones. This type of

Fig. 20. Cartoon showing the effects of facets to the growth morphology of directionally solidified products. A) one facet leads to a lamellar fabric; epitaxial twin planes allow for subsequent directions which can result in a plaited structure. B) Two facets restrict growth in lateral directions and lead to columnar shapes.

C) Three facets constrain growth in the third dimension, and deflects the maximum growth direction from that of maximum heat flow. C) lack of constitutional supercooling eliminate directional growth and three or more facets lead to porphyritic products with maximum elongation parallel to the isotherms (laminar textures).



growth corresponds to the open spherulitic growth of Miller (1977) and implies a second step in the direction of increasing growth anisotropy with decreasing supercooling. At least two facets are required to prevent dendrites from growing together in a transverse direction.

Growth is constrained in the third dimension when three facets form, and the maximum growth becomes directed away from the maximum heat flow direction. (Fig. 20C). The maximum growth is now pointed towards the nearest favorable crystallographic direction, as found for faceted dendritic growth in binary systems by Morris & Winegaard (1969).

The morphology is determined by the relative growth rate of the slower growing faces only when supercooling is almost totally eliminated. Under these conditions the temperatures are above the roughening temperature for all major facet planes. Maximum elongation is then parallel to the crystallographic *c*-direction in feldspars and parallel to the 010 plane, forming rhomb-shaped crystals typical of the lardalite interior (Fig. 20D). These rhomb-shaped feldspars are dominated by crystal faces like $\bar{2}01$, 110, & 110 (Ofstedahl, 1948), and growth rates for the various crystallographic directions can be ranked as $c < a < b$.

Pyroxenes

The most pronounced growth feature of the dendritic pyroxenes is clearly their curved dendritic shape. From the previous discussion it is clear that isolated dendrite stems imply faceted growth. Furthermore, the deviation of the faster growth direction from that normal to the cooling front implies that at least three facet directions control the growth. Taking the commoner facets of equilibrium shaped Ca-pyroxene as potential *f*-faces (Hartmann, 1982), it is likely that Ca-pyroxene forms facets in a manner slightly analogous to that of monoclinic ternary feldspars. That is, maximum density planes like 111 forms facets most readily following Gilmer (1977), while the loosely packed faces like 100 and 010 are late, and with the 001 and 110 planes as intermediate facets. Due to the competitive growth, the more rough interfaces 100 and 010 grow fast and tend to become oriented towards the maximum heat flow. Since these directions, however, are intercepted by the readily faceting 111 plane, maximum growth will become deflected at this intercept (Fig. 21). The growth thus produces an oblique directed dendrite with a planar (faceted) upper surface and a more irregular lower surface; the latter may produce secondary facets along 001 and 110 as the interdendritic growth proceeds.

Fig. 21. Crystallization of deflected pyroxene dendrites. Intersection between the fast growing 100 and 010 directions and a facet plane (e.g. (111)) results in a deflection of the maximum elongation of the dendrite from the direction of maximum heat flow.

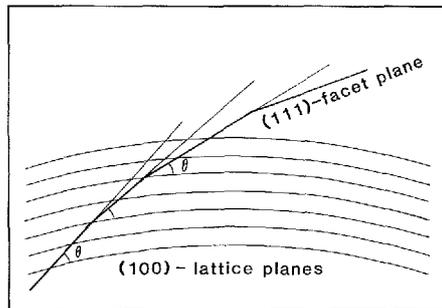
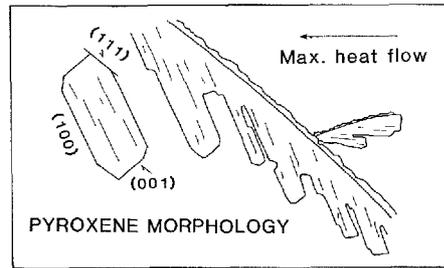


Fig. 22. Curved, external facets of pyroxene dendrites are the result of the interference between lattice-plane curvature of the fastest growing 100 direction, and the constraining facet plane 111, which gradually moves across the lattice curvature during growth and thereby leads to a curved overall shape. A curved dendritic shape thus, is the result of crystallographic factors alone.

The consistent curvature has been suggested to derive from flow of nourishing liquids past the crystallization front, imposing a compositional gradient (Lofgren & Donaldson, 1975). As realized by these authors, however, curvature is often developed in systems where flow is absent, and therefore a more general explanation is clearly wanted. The conspicuously curved lattice revealed by curved albite twins and deflected optic axes across single subgrains in the 010 plane of dendritic feldspars may contain the clue to an explanation of this phenomena.

Since the curvature along 100 is an overall lattice feature, the facet along 111 will gradually transgress upon this plane (Fig. 22). The angle θ between 100 and 111 however is a lattice constant and therefore as the transgression proceeds across the 100 plane, the 111 facet will become increasingly deflected. This leads to a systematic curvature of the resulting dendrite. Although curved dendritic growth of the analogous feldspars are more uncommon, a few examples (Fig. 23) show that the explanation also holds for feldspar. In Fig. 23 the curved nature of the 010 plane is seen from the weak differential coloring across the dendrite stem.

The above mechanism causes the pyroxenes to grow at an increasingly lower angle to maximum heat flow. When the angle of growth becomes sufficiently low, repeated nucleation occurs on the upper 111 facet, and a

new pyroxene starts growing initially perpendicular to the heat flow, in order to keep up with the continuous cooling; isotherms moving parallel to the contact. This process explains the regular, repeated nucleation seen in all the dendritic pyroxene zones (Fig. 24).

In summary, the formation of facets imposes a major control on the morphology of growing phases in supercooled conditions. This control is apparent in the sequence of crystal morphologies in the contact zones of the lardalite intrusion. The sequence of morphological types in these zones corresponds closely with that theoretically developed in a decreasing gradient of supercooling, which in turn causes a corresponding, stepwise increase in growth anisotropy (Fig. 25).

3.2 Compositional variations

Dendrite spacing

The regular spacing of dendrites in the inner dendrite zones of the lardalite intrusion, indicates steady-state growth at constant velocity. This view applies to both the open dendrite pyroxene zones and to the columnar feldspar dendrite zones. The spacing can be partly analyzed by assuming a diffusive boundary layer around each dendrite tip (Petersen, 1985). This layer has a radius of the characteristic diffusion distance; the problem of overlapping diffusion spheres is solved quantitatively in a fashion analogous to that of lamellar eutectic growth (Jackson & Hunt, 1966; Scherer & Uhlmann, 1975). When the interdendritic spacing is equal to twice the characteristic diffusion distance (Fig. 26A), lateral growth is inhibited by the equilibrium melt composition at the crystal melt interface. This is because excess solute from the interdendritic space

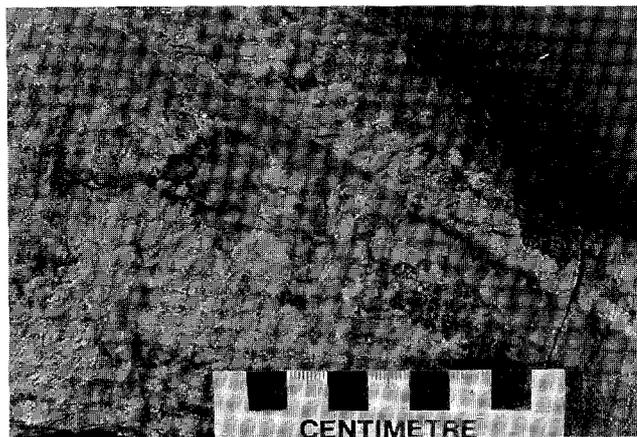


Fig. 23. Dendritic feldspar showing curved growth morphology. Convex, lattice planes can be recognized from the weakly discolored, transverse pattern on the feldspar stem. The crystal measures 18 cm. Secondary nucleation of pyroxene and nepheline on the upper surface tends to readjust growth towards the maximum heat flow.

Fig. 24. Sketch of curved pyroxenes showing repeated nucleation when growth is directed almost parallel to heat-flow. Compare with Fig. 13 and 23.

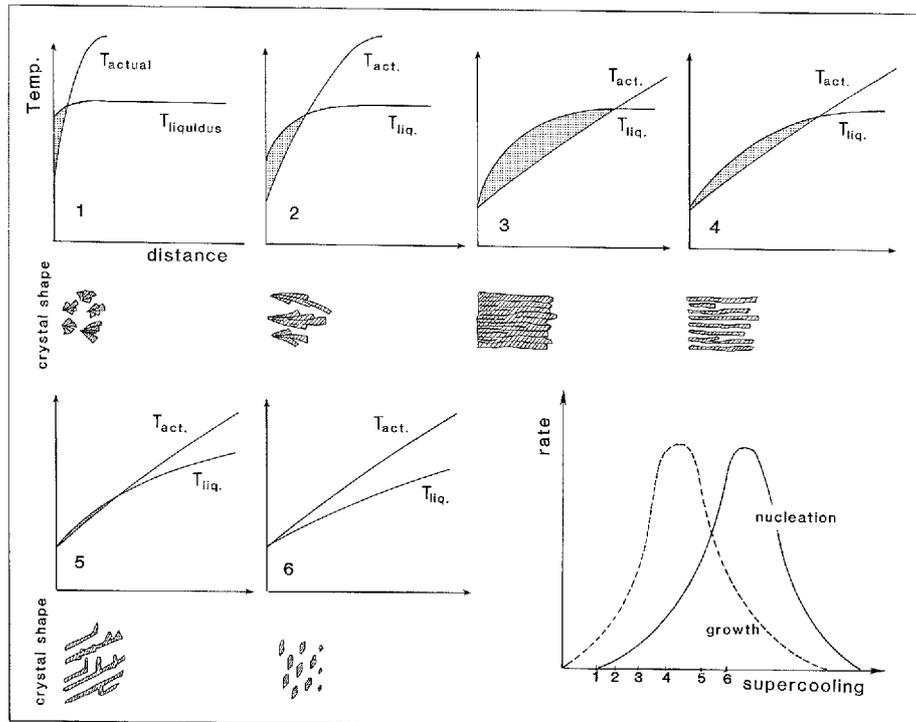
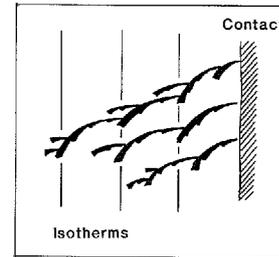


Fig. 25. Summary of growth morphologies and growth conditions for the contact zone feldspars in terms of thermal and constitutional supercooling from the margin (1) and inwards (2-6) (from Petersen, 1985). Initially high thermal supercooling result in semispherulitic crystals (1); increasing constitutional supercooling (2-3) results in directional growth. Formation of successive facets controls growth morphologies in (3-5) as thermal and constitutional supercooling decrease. In (6) the supercooling is eliminated and the crystals grow essentially parallel to the isotherms.

can be removed only in directions normal to the boundary layer. In this case, growth of the dendrite is constrained in all but one direction, and therefore, conformable to plane-front solidification.

When the interdendritic spacing is larger than twice the diffusion distance (Fig. 26B), transverse diffusion allows some lateral growth until mutually opposed diffusion spheres meet. Subsequent branching of dendrites may occur so as to adjust the solidification rate to a maximum for a given supersaturation (Flemings, 1974). In both cases, however, the process ultimately leads to unidirectional growth which is qualitatively comparable to plane-front solidification. Finally, if dendrite spacing is less than twice the diffusion distance (Fig. 26C), competitive solute enrichment occurs due to overlapping diffusion spheres. The growth of some dendrites is hindered and lags behind that of neighbouring dendrites. The situation then resembles that of case-B, where lateral growth or subsequent branching gradually eliminates transverse diffusion.

The above dendrite-spacing-regulation mechanism tends to adjust the dendrite spacing to optimum configuration for a given growth rate. Experimental evidence verifies that the spacing in cellular dendritic growth correlates with the rate of cooling rather than with the rate of motion of the isotherms (Chalmers, 1964, p.170). This implies that growth is controlled primarily by diffusion rather than by heat transfer. Dendrite spacing is therefore, governed by the characteristic thickness of the diffusion zone around the growing dendrite as argued previously by Howarth & Mondolfo (1962).

A systematic overall decrease in the feldspar to matrix ratio is found

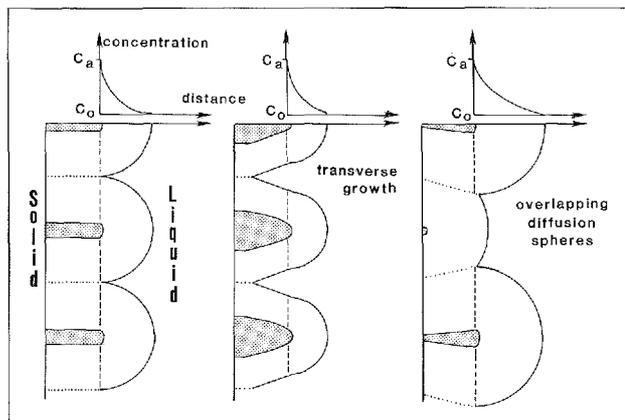


Fig. 26. Dendrite spacing regulation. Each dendrite tip acts as a source of solute, which build up in semi-spheric diffusion zones. (A) When the dendrite spacing is twice the characteristic diffusion zone, transverse growth cannot occur. In (B) is transverse growth possible until opposite diffusion spheres meet. In (C) overlapping diffusion spheres prevent growth of some dendrite stems.

from the pluton margin inwards. This suggests that the effective diffusion spheres increase in width inwards with decreasing cooling rate. Since a steady-state solidification is implied, solidification rate was diffusion controlled and the increasing diffusion spheres therefore not a result of decreasing solidification rate but rather the effect of increasing boundary layer width and less efficient removal of solute. This can happen because the convective flow velocity decreases with time or because the static diffusion zone widens as a result of increasing liquid viscosity.

Steady-state growth

Plane-front solidification has important implications for columnar-dendritic solidification if the maximum growth direction is not controlled by the interface reaction. As noted above, the solute content can be viewed as the amount of interdendritic material, which remains constant during steady-state solidification. Any change in the solidification rate, however, may result in compositional fluctuations about this constant value (Tiller et al., 1953). A change to slower growth rate momentarily reduces the solute accumulation rate and causes a temporary reduction in the solute diffusion zone. If additional branching is not developed, this reduction allows increased transverse growth of the pre-existing dendrite columns and creates a higher dendrite to matrix ratio in the growth front. The solid product will momentarily attain a lower concentration of solute (matrix) than the average melt. This imbalance results in a subsequent accumulation of solute in the solid phase until C_s again approaches steady-state composition.

For example, a steady-state dendritic solidification affected by a minor change in growth-rate from Y_1 to a slower growth rate Y_2 (Fig. 27) is accompanied by a reduction in solute accumulation. A momentary increase in dendrite width occurs, and causes the feldspar to occupy a larger volume proportion. Solute concentration in the solid product, viewed as the amount of interdendritic material along the solidification front is thereby lowered. Compared to the source liquid, this lower average solid composition renews enrichment of the solute to the boundary layer and in turn affects the dendrite to matrix ratio, which subsequently returns to the steady-state value.

The textural product of the above process may be a transverse, within-layer banding observed in some dendrite feldspar zones (Type-2 layering) (Fig. 2B). In these zones, numerous feldspar-enriched bands occur across the regularly spaced dendrite arrays and average compositions fluctuate about a constant composition.

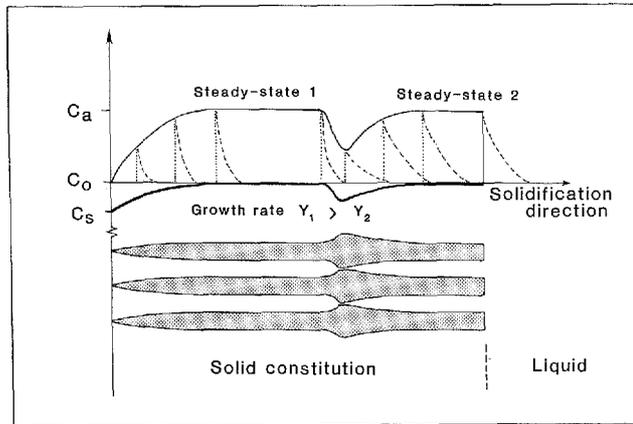


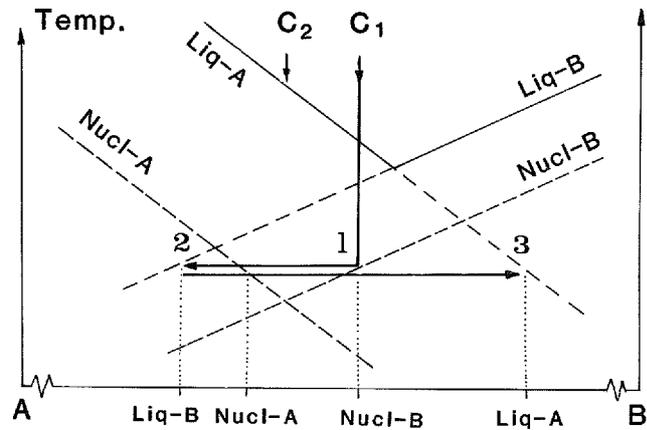
Fig. 27. Steady-state dendritic solidification with periodic changes in growth rate results in compositional fluctuation about constant composition. A sudden change from growth rate Y_1 to a slower rate Y_2 causes momentary reduction in solute content and thus a commensurate dendrite widening. This causes imbalance in the solute production which gradually forces crystallization to return to steady-state.

Complementary phase layering

Layering produced by the successive appearance of complementary phases is much more conspicuous than banding produced by slight variations in modal proportions resulting from steady-state growth of a single phase. Individual layers and intervening zones of open-dendritic or laminar-textured rocks with complementary phase layering typically have constant composition, but they may vary from polyphase dendritic to almost monomineralic. Despite their constant composition which indicates steady-state conditions these layers are not representative of bulk melt composition. Characteristics of damped oscillatory reactions are evident in the periodic nature of this layering. The nature of this periodicity suggests that complementary phase layering originates by competitive, oscillatory chemical processes. These processes may follow a nucleation-diffusion model initially proposed by Harker (1909) and recently reemphasized (McBirney & Noyes 1979), which in ideal terms, stems from the interaction of two parameters of growth with different rate properties such as e.g. heat flow and chemical diffusion.

As solidification moves from an arbitrary position x_1 to x_2 , heat flow varies linearly with time and elemental diffusion varies with the square of time (McBirney & Noyes, 1979). Therefore, at the start, the depletion of constituent elements moves faster than the cooling front and prevents subsequent nucleation. When the depletion zone is ultimately overtaken by the nucleation isotherm, a new burst of nucleation and subsequent growth will occur and create a repeated appearance of single phases in a rhythmically layered, solid product.

Fig. 28. (A) Complementary phase layering as a result of competitive nucleation and growth. See text for discussion.



Although intuitively satisfying there are indications that this process, at least for single phase nucleation, may reach steady-state and not oscillate (Allègre et al., 1981), unless it is triggered by an external mechanism such as the nucleation of a secondary phase. Steady-state solidification apparently does occur in the contact zones of the lardalite intrusion, since compositional variations are minor throughout most of the individual layers except for the occasional fluctuations discussed above and for conditions near the phase change, where gradual changes occur. Therefore a model is needed which allows the formation of even monomineralic products at a steady-state to become abruptly terminated and taken over by subsequent phases.

Solidification involving two complementary phases is discussed by using a simplified binary diagram (Fig. 28). The phases are stable below their respective liquidus curves, but only nucleate on the nucleation curves, which mark the necessary supercooling for stable nuclei of phase A and B respectively. In supercooled condition, the phases are stable below the extension of their respective liquidus lines.

A liquid of composition C_1 is supercooled until phase B nucleates and starts growing (Fig. 28). Assuming a broadly isothermal solidification, where latent heat of crystallization is removed through the solid phases, growth of phase-B drives the interface melt towards the left in the diagram. Depending on the size of the effective partition coefficient ($*K_D$), and the width of the static boundary layer, the composition of the solid may be very different or very similar to that of the source liquid. In the first case, the boundary-layer liquid will reach the nucleation curve for

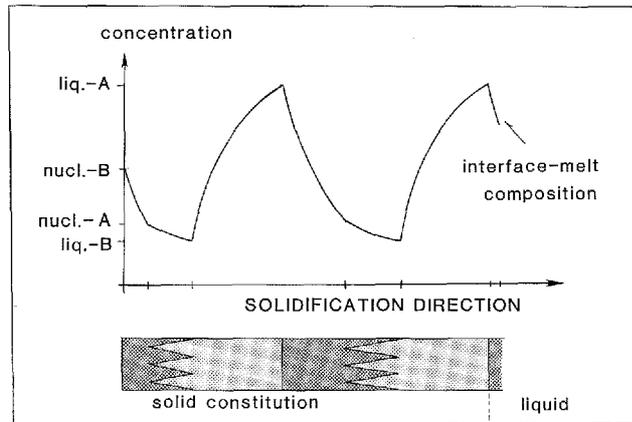
phase A rapidly and cause coprecipitation of A. In the latter case, steady-state solidification may develop, and remain until the reservoir liquid becomes saturated with respect to phase A, due to a decrease in temperature.

If phase A has a higher growth rate than B, phase A will take over the crystallization front, since the faster growing phase will enclose grains nuclei of the slower growing phase along the solidification front and development of phase B becomes prevented in the solid. Solute redistribution during this transitional stage is complex, but after completion it conforms again with the diffusion-sphere model discussed above, and thus a planar-front solidification model. From this stage (Fig. 28(2)), B-components now accumulate in the boundary layer relative to A-components and drive its composition towards the right.

When the interface liquid passes (1), nucleation of B will occur together with the continuous growth of A, but because of the growth rate differences discussed above, this rarely affects the boundary layer composition. In ideal isothermal solidification, the boundary layer proceeds directly towards the liquidus composition at (3). When it reaches the liquidus composition, equilibrium crystallization proceeds without any trend reversals until the source melt ultimately becomes differentiated. However, if solidification is not entirely isothermal, because latent heat of crystallization is not removed effectively from the faster growing A-dendrites or the liquidus is lowered because of volatile build up, then the interface temperature may rise above the liquidus temperature at (3) and promptly prevents further growth of phase-A. Nucleation and growth of the secondary phase-B, which is still in its stability field, then substantially influences the boundary layer and reverses the compositional trend. In contrast to the previous case, this phase contact will be sharp following the isotherms, and starting a new cycle of compositional layering.

An example of this process is sketched in Fig. 29, where A may represent feldspar or nepheline and B olivine or pyroxene. For a melt composition corresponding to C_1 in Fig. 28, the layered sequence would be (1) olivine (or/pyroxene), (2) olivine (/pyroxene) *and* feldspar (/nepheline), and (3) feldspar (/nepheline), assuming a higher growth rate for feldspar (/nepheline) than for olivine (/pyroxene). This sequence duplicates the order of appearance and layering observed in the marginal picrite zone of the lardalite intrusion towards the dendritic feldspar zone as well as the nepheline – pyroxene layering of the innermost dendrite zones. A slightly different composition at C_2 (in Fig. 28) causes primary precipitation of

Fig. 29. Schematic relations between two complementary phases like feldspar (A) and pyroxene (B); the former is assumed to have a higher growth rate. The upper boundaries for the faster growing phase are sharp whereas the upper boundaries for the slower growing phase are gradational. Phase-relations from Fig. 28.



feldspar, which would persist to produce a monomineralic feldspar zone. The feldspar continues to grow until heat-production brings the interface temperature above the liquidus. Numerous minor inclusions of Fe-Ti oxide dendrites in parts of the outermost, columnar feldspar zone (see Fig. 6 in Petersen, 1985) may represent engulfed nuclei of co-precipitating B-phases.

4 Equigranular solidification

4.1 Introduction

In the previous sections, solidification features related to the solid-liquid transformation with a fairly simple geometry, a unilaterally propagating planar front, were emphasized. The following sections examine aspects of solidification with a non-planar front and repeated nucleation. Similar compositional variations in comb-textured and cumulus-textured rocks suggest that the mechanisms of layering are related. The gradual transition from columnar to equigranular texture with decreasing supercooling suggests that both of these textures result from nucleation and growth in a similar heat-flow environment, but with different solidification rates and mechanisms, rather than being products of completely different solidification conditions. The purpose of this section is to examine the relationship between lateral solidification and compositional variations in equigranular rocks, and to discuss some petrological consequences.

The term, equigranular, implies equal grain size and equidimensional

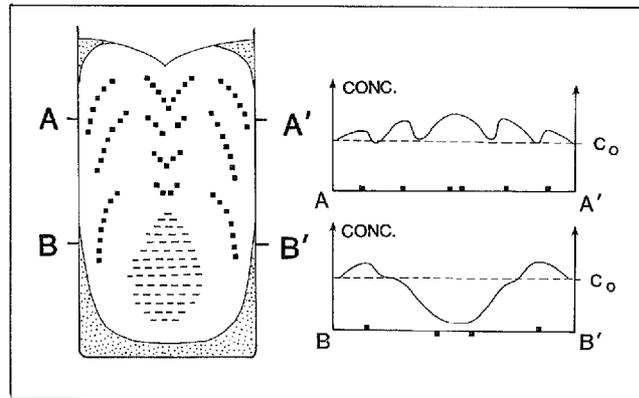
shapes of the constituent minerals. Its usage here is not completely accurate as the types of solidification structure with repeated nucleation discussed below may equally well produce any type of cumulus texture. The term is used, however, because it indicates an isotropic, non-directional texture as opposed to the strongly anisotropic, columnar or comb-layered products of lateral solidification discussed earlier. It is synonymous with the term, equiaxed texture, which is used in the metallurgical literature for analogous texture patterns (Chalmers, 1964).

Equigranular solidification textures may form in a static melt when the nucleation rate is sufficiently high to cause repeated nucleation along the cooling front rather than growth of preexisting phases. This process occurs in chilled margins where the nucleation rate is maximized at high supercooling, and the growth rate is small or moderate (see Fig. 25). Equigranular textures, however, may also form at *low* supercooling, when the maximum growth direction is constrained by crystallographic factors, and directed at some angle to the direction of maximum heat flow. Nucleation of new grains on pre-existing crystals subsequently occurs so as to adjust solidification rate to keep up with the conductive heat loss through the wall rock. Despite their similar texture patterns, the conditions for the formation of the above two equigranular products are fundamentally different, and these differences become reflected in the contrasting compositional properties of the products.

As for directional solidification of dendritic and cellular products, important contributions to our understanding of unilateral solidification of granular textured products is provided by experimental studies of continuous casting processes in axisymmetrical ingots. These experiments show that systematic compositional variations occur in almost any solidification geometry (Flemings, 1974; p245). This type of variation, known to metallographers as macrosegregation, is derived mainly from the interaction of diffusive boundary layers with fluid flow in the interdendritic region, also called the mushy zone. Other mechanisms such as floating or settling of crystals or convective flow of liquid plus solid may contribute to the macrosegregation in special cases (Mehrabian, 1984).

Plutonic intrusions in a simplified sense are vast castings of silicate magmas, and are analogous to the much smaller scale metallurgical castings in which macrosegregation is important. It seems fitting therefore to examine the origin of compositional variations in plutons with respect to crystallization kinetics and macrosegregation. Compositional variations which are inconsistent with the variations expected along the liquid line of descent of the magma are particularly significant in this respect.

Fig. 30. Macrosegregation in typical cast ingot. Square dots mark positive segregation whereas horizontal stipple is negative segregation. Chilled products are finely dotted. Right diagrams show idealized compositional variations across two sections A-A' and B-B'. The axisymmetrical distribution is pronounced; positive segregation fluctuates in the upper portion of the ingot whereas negative segregation dominates the lower portion.



Macrosegregation is usually expressed as the relative deviation from average composition of the source melt; positive segregation corresponds to a more differentiated state while negative segregation represents a less evolved composition (Flemings, 1974). Axisymmetrical ingots often exhibit substantial segregation in the vertical as well as the horizontal direction. A typical example of macrosegregation in cast metal ingots illustrates this qualitatively: In the upper part of the ingot (cross section A-A', Fig. 30), several zones of positive segregation are separated by neutral zones and form a regular compositional variation which fluctuates through the ingot. This variation resembles the regular trend reversals in cryptic layering observed in many igneous bodies, e.g. the Stillwater Intrusion (Jackson, 1961). In a lower part of the same ingot (cross section B-B', Fig. 30), compositional variations are expressed by an outer margin of positive segregation and a core of substantial negative segregation. Geologically, this trend compares with the reverse compositional variation as found e.g. in the lardalite pluton (Petersen, 1978b). It is also comparable to the overall variation found in many anorthosite massifs, where differentiated noritic rocks form a marginal facies and more primitive anorthosites generally make up a central core (Michot & Michot, 1969; Ashwal & Seifert, 1980; Demaiffe & Hertogen, 1981).

4.2 The chain mechanism of heterogeneous nucleation

A fundamental prerequisite for solidification is the nucleation of solid phases which occurs by the condensation of sufficiently large clusters of

constituent molecules or polymers within the melt. Nucleation is classified as either heterogeneous or homogeneous, depending on whether it occurs with or without the aid of foreign substrates (Carmichael et al., 1974; Kirkpatrick, 1975; 1983). The stability of newly formed nuclei depends on their surface free energy, which is often expressed in terms of the critical size or critical radius for the nucleus at a given amount of supercooling (Shewmon, 1969). Thus, homogeneous nucleation requires a cluster of a certain size to allow subsequent growth. Below this size the nucleus will redissolve and become extinct. The tendency for heterogeneous nucleation of a particular phase can be expressed by its 'wetting' factor, which equals the cosine of the equilibrium contact angle θ (Fig. 31). The smaller this contact angle is, the smaller the volume fraction of a critical nucleus sphere will be, which produces a stable surface curvature that allows subsequent growth (Chalmers, 1964, p. 77; Shewmon, 1969, p. 160). Structural compatibility between the nucleus and the substrate surface reduces the contact angle and increases the chance for heterogeneous nucleation.

For these reasons, it is likely that all nucleation in natural magmas is heterogeneous and occurs preferentially on analogous materials (Kirkpatrick, 1975). Experimental studies of the solidification properties of natural silicate melts have strongly supported this view (Lofgren, 1983).

In a lateral cooling configuration as described above, it is probable that nucleation and growth preferentially occur on pre-existing crystals along the borders of the magma. Growth is facilitated in crystals that are in mutual contact and attached to the wall rock, as compared to freely suspended crystals, since the latent heat of crystallization can be extracted through the solid media. Any internal nucleus will eventually be overgrown and eliminated as solidification proceeds (Fig. 32). Convection in the magma interior may cause transport of early detached crystals along the magma chamber borders (Irvine, 1979). However, when transferred

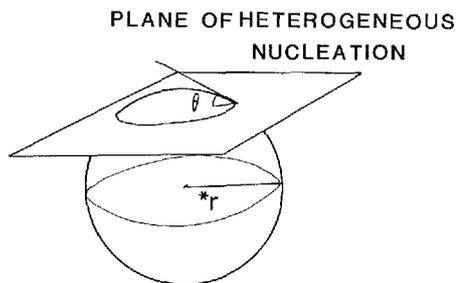
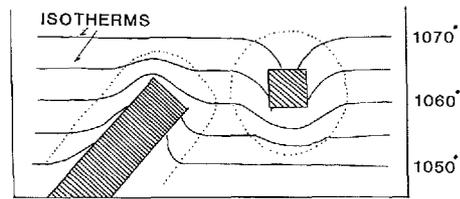


Fig. 31 The portion of an idealized, critical nucleus with radius r^* above the plane of heterogeneous nucleation has a stable surface curvature that allows subsequent growth. The critical contact angle marks the maximum curvature of a stable surface, and thus the fraction of a critical nucleus that is necessary for heterogeneous nucleation.

Fig. 32. Growth conditions for freely suspended versus contact-attached crystals. Heat of crystallization is build up around the freely growing crystal whereas heat can be retracted through the stem of the attached crystal. The thermal distribution thus favours growth of attached crystals over suspended individuals.



towards the hotter interior, newly formed crystals will generally dissolve, and the interior may be considered free of nuclei and fairly homogeneous within each convective unit. This applies also in the case of stratified magmas which are increasingly being accepted in the petrological literature (e.g. Hildreth, 1979; Irvine, 1980; Huppert & Turner, 1981).

When growth of a solid phase is directed away from the heat-flow gradient, supercooling at the solidification front gradually increases because the isotherms continue to move towards the magma center at a linear rate normal to the gradient. Near the crystal interface, the isotherms are slightly deflected locally, but since the growth rate is much slower than heat-flow, this will not affect the overall increasing separation of the isotherms and the solid interface. When this separation reaches a critical value, heterogeneous nucleation will occur on the crystal face rather than continued growth, in order to achieve a higher solidification rate. This grain multiplication process is excellently illustrated by the curved dendritic pyroxenes, described in section 2; when the growth of these pyroxenes, because of the continuous curvature, becomes almost parallel to the isotherms repeated nucleation occurs on the upper surface of the dendrite (Fig. 24).

The formation of most equigranular products occurs at near-equilibrium conditions where growth anisotropy is high and solidification is interface controlled (Kirkpatrick, 1975; Gilmer, 1977). Under these conditions, the growth direction is determined by crystallographic directions of the nucleus rather than by heat flow or chemical diffusion. Low-angle branching that adjusts the growth direction at high supercooling does not occur. The maximum growth direction is therefore determined by the crystallographic orientation of the heterogeneous nucleus. The competition between maximum growth direction and maximum heat-flow direction then results in a repeated nucleation and growth process that leads to the formation of equigranular and cumulus textured pro-

ducts with a lateral solidification geometry. As the growth of a phase enriches the liquid in solute, subsequent heterogeneous nucleation of a secondary phase will likely occur on the surface of the first, where solute enrichment is highest. With the subsequent depletion of solute, the growth of the first phase is enhanced, and the process therefore leads to clustered solid aggregates.

Textural evidence showing such clustered nucleation-growth aggregates (bundle texture), is suggested to be a common, yet poorly recognized solidification structure in cumulate rocks (Campbell, 1982). The occurrence of bundle textures offers a reasonable explanation for the formation of large blocks of solidified material, e.g. along the interface of a hot, primitive magma that flows past a cooler melt, that may eventually become big enough to overcome the density-contrast, yield-strength barrier discussed by McBirney & Noyes (1979, Fig. 3), and settle through the liquid. However, it can also be the most important solidification mechanism along the cooler contacts of a magma chamber, causing crystallization to occur along a well defined interface between liquid and solid.

Preliminary experimental studies of solidification and macrosegregation of silicate melts in crucibles clearly emphasize the importance of the cluster or chain nucleation. Fig. 33 shows a two-dimensional expression of the solid-liquid interface produced with moderate supercooling. Depending on the exact cooling rate, the interface attains a regular planar or a highly irregular, indented shape. In three dimensions, the indented shape of this type of interface allows liquid to flow through the intergranular space in a manner that is analogous to interdendritic melt percolation in metallic castings (Flemings, 1974). These similarities justify the application of macrosegregation modeling in metals to silicate magmas.

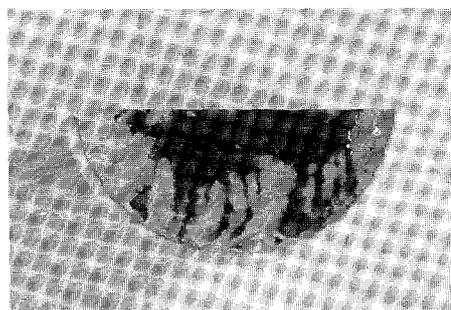


Fig. 33. Pt-plated, ceramic crucible with 2:1 plagioclase:diopside melt cooled from 25° superheat at moderate rate. Repeated nucleation on preexisting crystals result in crystalline trains propagating towards the center. Internal nucleation is extremely rare.

4.3 Compositional variations

When a solid phase crystallizes, solute is rejected from the crystal-liquid interface. The amount of this rejection is governed by the partition coefficient at equilibrium conditions (McIntyre, 1963). At supercooled conditions, crystallization may occur isothermally, and the solid approaches the composition given by the solidus at this temperature until the continued enrichment of solute causes the melt to reach liquidus composition at C_o' (Fig. 34A). Subsequent crystallization then occurs at equilibrium conditions. When the crystallization rate is sufficiently small, the rejection of solute ultimately affects the residual liquid and gradually changes its composition. Progressive crystallization under these conditions results in a continuous enrichment of solute in the solid, as the source liquid becomes increasingly fractionated (Fig. 34B).

In the solidification of melts where the growth rate exceeds diffusion, the concentration of solute will increase near the solid-liquid interface (Fig. 34C). When a sufficiently large solute concentration has been established, one of two things may happen: diffusion of solute away from the boundary layer will keep up with the solidification rate and solidification becomes steady-state (Tiller et al., 1953; Henderson & Williams, 1977); or, the solute enrichment will reach saturation with respect to a second, near-liquidus phase (C_B in Fig. 34C).

Because the diffusion rate diminishes with the square of distance from the interface (Hoffmann, 1980), and the solidification rate at low supercooling varies linearly with time (Kirkpatrick, 1975), it is unlikely that growth attains steady-state in multicomponent systems before solute saturation occurs. Development of mono- or biminerals are thus generally prevented in a static melt solidification. However, when fluid flow affects a portion of the diffusive boundary layer, solute is removed at much faster rate. The conditions for a balance of input (solute production during crystallization) and output (diffusion plus removal by fluid flow) are thereby readily established, and particularly at a low absolute solute enrichment are conditions for steady-state solidification readily attained (Fig. 34D). In a closed system solute enrichment ultimately alters the source liquid and a final compositional transient of residual liquid forms which is highly fractionated (Fig. 34E). When this fraction crystallizes, the solid shows an appropriate increase in solute content.

Solute removal can be enhanced by the overlap of thermal convection in the magma interior with the diffusive boundary layer. Solute which diffuses into the convective unit cannot exceed its concentration at the

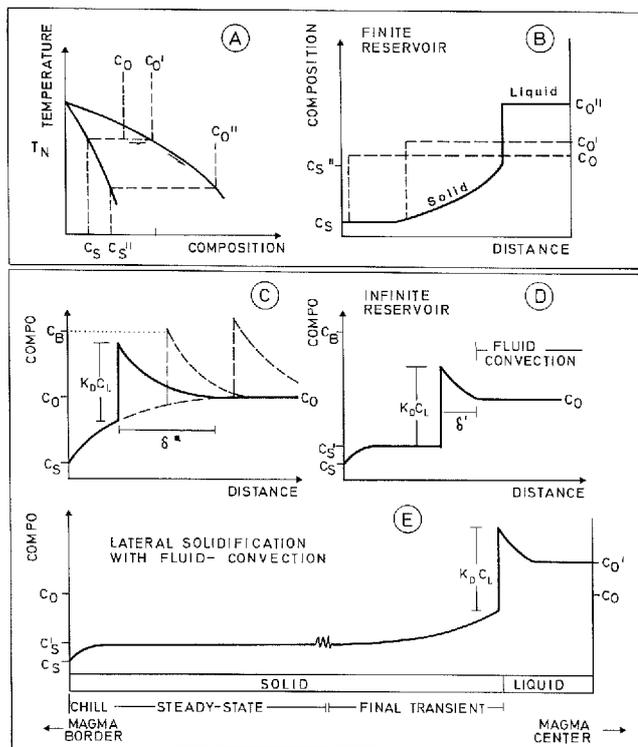
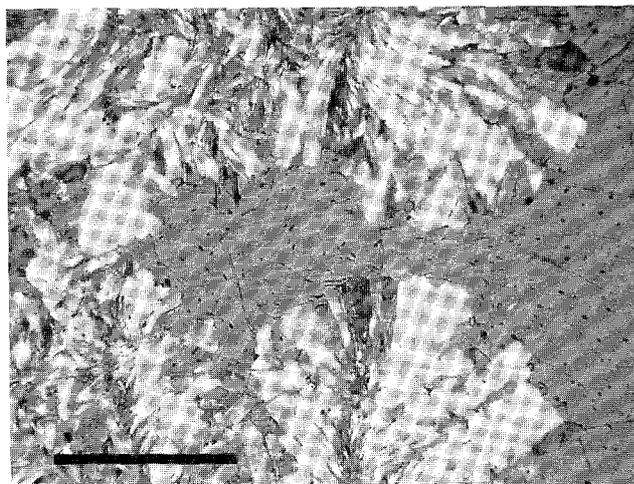


Fig. 34. Solute redistribution during plane-front solidification conditions. See text for details.

flow interface. The width of the remaining, static portion of the boundary layer δ' , then determines the maximum amount of solute enrichment that develops for a given diffusion coefficient (Fig. 34D). The effect of this process can be quantified by using a different effective diffusion coefficient in the solute redistribution equations (Flemings, 1974). Thermal and rheologic properties of natural silicate magmas support the presence of fluid convection (Bartlett, 1969). Such magmas may solidify at steady-state under broadly isothermal conditions and produce nearly constant compositions after an initial transient.

A second type of fluid flow, percolating flow between growing bundles of solid phases, results in substantial mass transfer through the mushy zone of crystallization (Flemings, 1974; Mehrabian, 1984). This flow is established by a considerable volume reduction during solidification at highly irregular solidification interfaces, and is typically encountered in the dendritic solidification of metallic melts. As shown in Fig. 35, the percolating flow, however, also plays a prominent role in comparatively viscous silicate melts upon progressive solidification, and should not be overlooked.

Fig. 35. *Experimental charge of ternary feldspars solidified by heterogeneous nucleation and growth from the capsule walls. Deformed vapor bubbles reveal fluid flow in the intercrystalline liquid as the result of volume reduction during prograde crystallization. This indicates that significant percolation flow exists in silicate melts upon solidification. Scale bar measures 1 mm.*



During steady-state solidification the solid composition is defined by the relation $C'_s = \star k_d C_o$, where C_o is the composition of the source liquid, and $\star k_d$ the effective partition coefficient at steady-state. The extent of steady-state crystallization depends on the differentiation of the residual liquid, and on the difference between the source liquid (C_o), and the solid product of the steady-state solidification (C'_s) (Fig. 34D). This difference is a reflection of the magnitude of the effective partition coefficient $\star k_D$. Values of $\star k_D$ near unity may produce substantial amounts of steady-state growth since the residual liquid is only slightly affected. In contrast, large or small $\star k_D$ -values will tend to reduce the amount of steady-state solidification since the source liquid becomes differentiated more rapidly.

The distribution of solute during lateral solidification can be modeled quantitatively and applied to polyphase systems by considering the solvent as the sum of constituent components in the solid product, and the solute as the sum of complementary components in the melt. Chemically, the characteristics of these components can be defined as a matrix of the type $B = M - A$, where M is the composition of the source melt; A is the composition of the solvent (= solid aggregate), and B the solute, readily calculated for various solid aggregates by e.g. petrographic mixing programmes (Wright & Doherty, 1970).

The bulk chemical modeling discussed above accounts for the overall variation of the solid product during solidification, and explains the formation of gradual changes in the initial or final transients, which may

or may not include an intervening steady-state solidification. Whether the size of the effective partition coefficient k_d is greater or less than one, the solid product may attain a prograde or a reverse, overall compositional zonation. Its application, however, is ideally restricted to a plane front solidification and simultaneous crystallization of the main constituent phases.

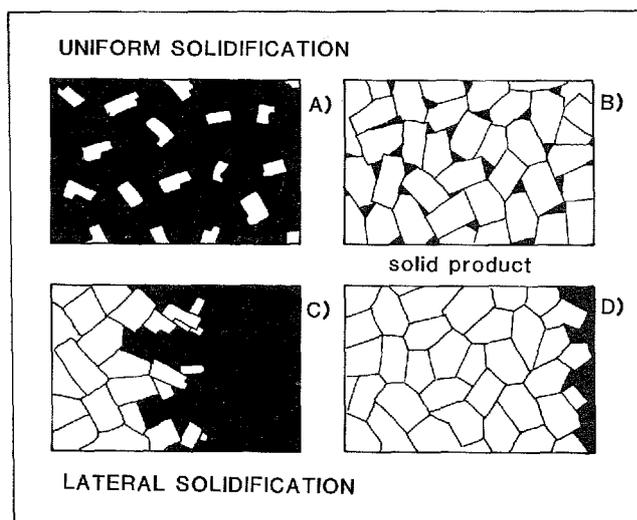
These conditions are readily met in eutectic crystallization where simultaneous precipitation occurs with a planar front (Flemings, 1974). The plane front is achieved because transverse diffusion and phase-spacing regulations adjust the bulk interface liquid to eutectic composition at any given growth rate, and the product solidifies at steady-state (Mollard & Flemings, 1967; Elliott, 1977). A plane front is also achieved at dendritic solidification because dendrite spacing adjusts to minimize transverse diffusion and the solute redistribution modeling becomes equivalent to that of eutectic solidification (Scherer & Uhlmann, 1975). In this case, secondary phases precipitate from the trapped interdendritic melt. Their crystallization, however, does not change the overall composition of the product. Therefore, in terms of solute redistribution, the dendritic solidification can be viewed as a special case of multicomponent solidification where one phase, because of a higher growth rate, is leading. In order to understand the more general situation, analysis of solidification geometries and kinetics of crystallization is therefore required.

4.4 *The solidification geometry*

Two extreme types of solidification geometries provide a useful basis for the discussion of directional chemical segregation. In one case, internal nucleation is abundant and approximately simultaneous within the entire melt fraction. Such nucleation causes spheroidal diffusion zones, enriched in residual solute, to form around each nucleus during subsequent growth (Fig. 36A-B). Because this solute enrichment occurs in an isotropic configuration, no compositional gradient exists between different portions of the melt, and the expelled inter-mineral solute cannot migrate by long-range diffusion. Solute content in the intercumulus melt therefore increases continuously with progressive crystallization and the composition of the solid phase assemblage becomes increasingly zoned and fractionated.

This type of solidification geometry, termed uniform, due to the uniform distribution of nucleation, corresponds to the traditional view on the origin of cumulate textures (Wager et al., 1960). Escape of variable

Fig. 36. Principal solidification geometries. A-B are the result of uniform nucleation and isotropic growth. C-D are formed by heterogeneous nucleation and subsequent lateral growth. The rejected solute accumulates ahead of the solidification front in D), contrary to the uniform crystallization geometry B), where solute is evenly trapped within the sample.



amounts of intercumulus liquid is required for fractionation, but the mechanisms of this segregation remain somewhat controversial. This type solidification may represent a special case, where thermal supercooling is very high, such as just after injection of the magma or when the melt is suffering from vigorous turbulent flow due to emplacement or volcanic eruption, but probably may not apply to the general solidification of resting magma chambers as often thought. Chilled margins are obvious natural examples of this type of solidification mechanism, and so are volcanic products and some porphyritic-textured, contact rocks.

In the other solidification geometry, designated here as lateral solidification, nucleation and growth project laterally from the contact zone towards the interior of the magma. This can occur by the chain nucleation and growth mechanism (Fig. 36C-D). The residual components accumulate in the boundary layer ahead of the solidification front and form a compositional gradient. This gradient creates conditions for the removal of solute by directional diffusion and fluid flow, and its ultimate accumulation in residual solidification pockets (Fig. 36D). Certain monomineralic, comb-layered rocks are obvious examples of this type of solidification (Petersen, 1985).

Either of the above solidification geometries lead to differentiation provided removal of intercumulus liquid occurs. This is often expressed by the amount of adcumulus growth (Wager et al., 1960; Wager & Brown, 1967) or by the proportion of mesostasis – pore liquid (Hender-

son, 1970). In the uniform model, the removal of interstitial solute requires an effective percolation of fluid through a densely nucleated crystal mush to produce fractionation. Since the presence of nuclei or crystalline particles increases the viscosity of the liquid substantially (Komar, 1972), this process becomes increasingly difficult. Removal of interstitial liquid by fluid flow possibly occurs during compaction by a process sometimes referred to as filter pressing, although, the capacity of this mechanism to explain enhanced adcumulus growth is limited (Bowen, 1928, p. 157; Emmons, 1940; Carmichael et al., 1974). Diffusion through the interstitial liquid is possible, but this requires the presence of an overall compositional gradient and it becomes almost prohibitively slow beyond distances on the order of the average grain size. Effective fractionation and adcumulus growth in the uniform solidification geometry is therefore difficult, and at best occurs at extremely slow solidification rates.

In contrast, the chain nucleation-growth model suggested above allows the continuous removal of excess solute during growth by diffusion into the boundary layer and by fluid flow along the solidification front. Depending upon the shape of the solidification front, planar or highly indented, this type of solidification creates a diffusive boundary layer which is more or less parallel to the isotherms, and causes compositional variations to follow a plane-front solute redistribution model. This process can therefore lead to the formation of cyclic compositional variations (type-2 layering); complementary phase layering (type-1 layering); steady-state growth of mono- or polycrystalline products, or simply gradual differentiation as was seen for the comb-layered products. The textures, however, will be those of repeated nucleation, i.e. cumulus textures instead of a columnar directional fabric. Another very important implication of the above textural considerations is that many of the planar, contact parallel structures found in igneous rocks are likely to reflect successive time interfaces during crystallization, i.e. solidification isochrons, which are parallel to the isotherms.

5. Applications of lateral solidification and macrosegregation to petrogenetic modeling

Many plutonic complexes show regular compositional variations which are readily explicable in terms of lateral solidification. In many zoned plutons, the variation, however, shows prograde differentiation from the margin towards the center, for example the Rocklin pluton (Swanson, 1978) and the Tuolumne Series (Batemann & Chappel, 1979) in Sierra Nevada, California, thus, a variation which can well be accounted for by traditional fractional crystallization. In some plutons compositional variations occur which are reverse to this trend; differentiation seems to project from the core towards the margin, and clearly present some problems with respect to traditional petrogenetic modelling.

A few examples of such 'anomalous' cases will be briefly considered here in order to illustrate advances of the macrosegregation concept over traditional petrogenetic modelling and to demonstrate the capacity of this hypothesis to account for both normal and reverse overall compositional variations as well as the formation of homogeneous or rhythmically layered solidification products.

The Hidra anorthosite-norite is a zoned intrusive body in the SW part of the Egersund anorthosite province, SW Norway (Demaiffe & Hetogen, 1981). It constitutes the youngest anorthosite body in the region and has suffered minimum post-emplacement textural modification such as granulation of feldspars and pyroxenes, contrary to most other anorthosite bodies of the province. The primary nature of this complex therefore makes its textural and compositional relations particularly significant.

The complex measures about 6×20 km and shows a regular concentric distribution of different rock facies. Following a narrow, fine grained and porphyritic, jotunitic border facies is a marginal, ophitic-textured norite with average grain size about 2 mm which grades into coarser norite and finally into coarse grained leuconorite and anorthosite with single feldspars measuring 5×15 mm and up to dm size in the core of the complex. No internal contacts between these facies have been observed.

Plagioclase occurs as cumulus feldspars with moderate, normal zonation from An_{50} to An_{45} in the central, anorthositic and leuconoritic part. The most Ca-rich feldspars from the ophitic-textured border zone, show compositions which are equivalent to the rim compositions of the inter-

ior plagioclases, about An_{45} , whereas interstitial feldspars in this zone rate about An_{35} and are additionally fairly Or-rich, about Or_{14} . These compositions accordingly suggest that the border-zone rocks are more evolved than the interior rock facies in this complex.

Ca-poor pyroxenes which constitute the dominant ferromagnesian phase in the Hydra complex, show similar relations. The interior, anorthosite hosted orthopyroxene of En_{70} forms interstitial grains, completely bounded by cumulus plagioclases. Approaching the marginal facies, both plagioclase and orthopyroxene form cumulus phases with pyroxene compositions about $En_{65}Fs_{33}Wo_2$, – thus more evolved than the interior ones. The ophitic-textured border norites show even more pronounced Fe-enrichment about $En_{57}Fs_{41}Wo_2$. Ca-rich pyroxenes are present in this zone as discrete grains, whereas elsewhere in the complex Ca-rich pyroxenes occur only as minor anhedral or interstitial grains. Biotite, apatite and interstitial K-feldspar, the latter showing a spectacular intercumulitic, symplectitic intergrowth with quartz occur exclusively in the border facies, and emphasize the more differentiated character of this part of the pluton.

In addition to the increasing mineral differentiation towards the margin, the Hydra complex also shows a significant modal increase in ferromagnesian phases from the core towards the margin, i.e. in the direction of prograde differentiation, a common feature in anorthosite massifs which led Ashwall (1982) to suggest that mafic and Fe-enriched rocks constitute residual products of anorthosite formation. Both the increasing differentiation towards the margin and the decreasing mafic content towards the center, however, are incompatible with the overall compositional variation in calc-alkaline or tholeiitic rock suites that crystallizes from the margin inwards.

The reverse compositional variation in the Hydra complex is conformable with inwards solidification assuming a decreasing cooling rate and thus increasing negative macrosegregation. Differentiated rocks formed in the initially steep thermal gradient at the margins, where removal of solute was minimum. Abundant nucleation in a thermally supercooled contact may form a chilled border facies. The ophitic textured, marginal facies of this pluton, however, are considered subsequent products of repeated nucleation and growth at subliquidus conditions, in a fairly steep thermal gradient. The rapid solidification resulted in local supersaturation and precipitation of accessory phases such as apatite and K-feldspar. Prolonged heating from the hotter interior allows solid diffusion to smoothen out the initial mineral zonations in the contact zone rocks.

Later, as the cooling rate decreases, increasing amounts of rejected solute are removed by diffusion and convection, and thus result in increasingly negative segregation. The reciprocal, residual liquid of this solidification process would become increasingly mafic in accordance with the findings of Ashwal (1982) and also accounts for the occurrence of minor, intrusive Fe-Ti ore bodies in the district.

A number of small intrusive, alkali-rich gabbroic stocks, less than 400 m wide, which additionally occur in the anorthosite province of south Norway, reveal interesting analogies in composition and textures with the ophitic-textured border facies of the Hydra massif. In contrast to the compositionally zoned anorthosite-norite complex, these minor gabbro plugs are exceedingly homogeneous and display no variation from the border towards the center. Instead, a remarkable zonation is given by the constituent minerals which support the textural relations in suggesting an enhanced, *in situ* fractionation.

In one of these stocks, the Lyngdal hyperite, plagioclases are lath shaped about 1×4 mm and display normal zonation from about An₄₅ to antiperthitic margins about An₃₅. Myrmekitic fringes on the antiperthitic margins project into interstitial orthoclase and quartz. This textural sequence is conformable with a closed-system fractionation with progressive build up of residual components in the trapped interstitial liquid. The formation of myrmekitic fringes may actually represent coordinated growth of a cotectic plagioclase-quartz pair which is succeeded by residual quartz or K-feldspar as indicated by the textural relations. Ca-poor pyroxenes are the predominant ferromagnesian minerals in the hyperite and span a fairly narrow range from about En₆₃ to En₅₉. Apatite is abundantly present and forms stubby euhedral grains.

Despite the immediate differences in texture and compositional variation between these two plutons, remarkable similarities exist. The range of feldspar and pyroxene compositions in the anorthosite complex is almost identical to that of the zoned minerals of the hyperite. The latter variation, however, projects from the center of single crystals into the interstitial space, while the former shows the same overall variation from the core of the complex to the margin, although, as mentioned, opposite to the expected trend of differentiation.

The contrasting textural evolution of these two plutons is summarized in Fig. 37. In a closed system, such as the marginal facies of the Hydra complex or the Lyngdal hyperite intrusion, early Ca-rich plagioclase becomes progressively more Na-rich when poikilitic pyroxene joins the assemblage (Fig. 37,1-2). Subsequently, plagioclases with antiperthitic

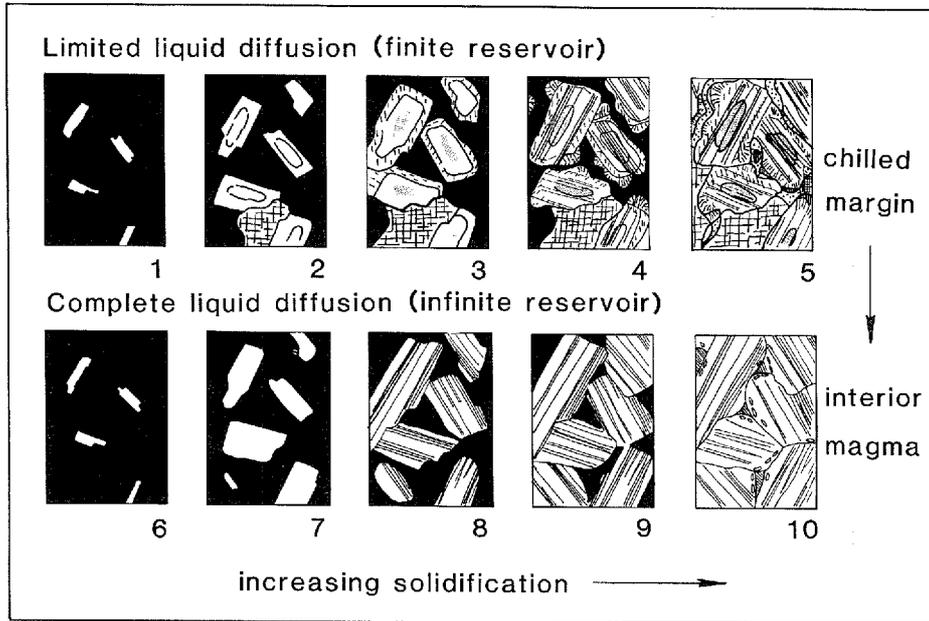


Fig. 37. Contrasting solid products as formed by different solidification conditions. 1-5 show the progressive fractionation in a uniform solidification geometry with no liquid diffusion. The solid product (5) equals the composition of the initial liquid (1). In 6-10, the initial liquid (6) is similar to that above. Removal of excess solute due to liquid diffusion and/or lateral solidification geometry, allows continued precipitation of the initial phase assemblage (7-8).

exsolution lamellae develop (3), which later grade into myrmekitic fringes (4) and ultimately into interstitial quartz and K-feldspar (5), which mark the closing compositions. Apatite may additionally form in local, supersaturated boundary liquids. This sequence completely mimics the textural variation observed in the Lyngdal hyperite.

The other extreme, illustrated by Fig. 37, 6-10, forms when solute is removed from the intercrystalline liquid. Depending on the solidification geometry this process can allow perfect adcumulus growth and continuous precipitation of the initial, high-temperature phases as long as the source liquid composition and the solidification conditions remain constant. A complete gradation from case 1-5 at the margin of the intrusive body to case 6-10 in the core can be expected in the course of crystallization and progressive cooling. In case of the anorthosite body, prolonged cooling at subliquidus temperatures may gradually eliminate initial mineral zonation in the marginal facies. Identical geochemical properties of the Lyngdal hyperite and the marginal facies of the Hidra

complex, including trace element data (Petersen, in prep.), show that the former may represent an appropriate parent magma for anorthosite.

In a similar case of competitive nucleation and growth, oscillating chemical variations in the boundary layer can relate to the formation of complementary phase layering. Two principal types of layering found in the predominantly homogeneous larvikite intrusions of the southern Oslo region emphasize this point. One type is graded, compositional phase layering, with sharp lower contacts, that is equivalent to Skaergaard type layering except for its nearly vertical orientation, which makes an origin by gravitative crystal settling difficult. The other type, which is slightly more common, is characterized by faint layering caused by the cyclic appearance of intercumulus, poikilitic pyroxenes. This variety, which is even more difficult to relate to a crystal settling origin, is explicable by the periodic nucleation of faster growing pyroxenes along isotherms that follow a nearly planar front. These examples show that the composition of the solid product depends not entirely on the compositional gradient in the boundary layer, but also on the kinetics of the crystallization, such as the relative nucleation and growth rates.

One of the important features of the repeated nucleation and lateral solidification suggested here, is that solidification occurs along isotherms, which are broadly parallel to the contacts. This will produce a largely concentric solidification pattern. Accordingly, any changes in the crystallization conditions, such as the formation of rhythmic or cryptic layering, are to be recorded along these concentric interfaces. Differences in the thermal gradients (cooling-rate), on the other hand, may cause the phase assemblages to change along the solidification front.

A possible example of this is shown in the Farsund charnockite, SW Norway which was injected adjacent to major anorthosite massifs of SW Rogaland, that produced a substantial thermal gradient in the country rocks as indicated by pronounced metamorphic aureoles. The concentric pattern of e.g. the FeO content in the charnockite (Fig. 38) is conformable with a solidification along isotherms from the margin inwards. The phase assemblages and compositional properties of the constituent minerals, however, vary laterally across the pluton and exhibit more primitive compositions in the southeastern, more rapidly cooled part, and higher temperature assemblages like fayalite instead of orthopyroxene in the northwest (Petersen, in prep.).

The interference between isotherms and cooling rate changes may particularly affect the composition of the solid products in a flat sheet of magma, where cooling rate at the extremities will be much higher than

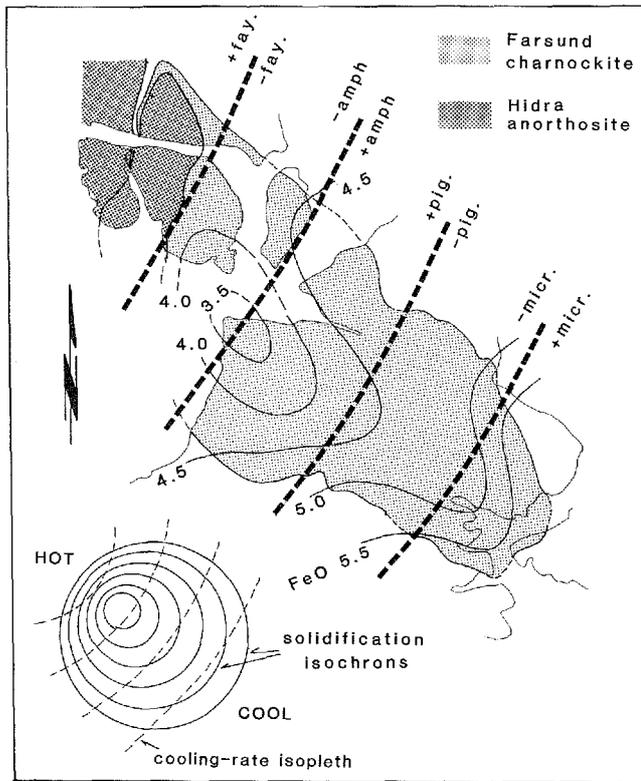
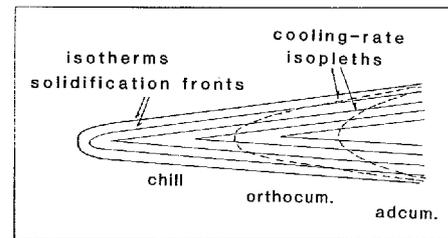


Fig. 38. Concentric patterns of isotherms and solidification fronts can be reflected by cryptic compositional layering such as the FeO-content in the Farsund charnockite. Cooling rate isopleths which influence the phase assemblages, however, can vary across the pluton according to a regional thermal gradient.

the central portion. Fig. 39 shows a cooling-rate isopleth of such a magma type following the data of Jaeger (1968). In the flat sheet, the isotherms will follow the boundaries, and the solidification interfaces therefore become practically horizontal. A pronounced, lateral, compositional variation along the solidification front can therefore develop due to changing cooling-rate and different diffusion efficiency; the rapidly cooled extremities will appear more differentiated because of higher amounts of trapped residual liquid here, whereas progressively slower cooled parts will appear less differentiated because of the increasing efficiency of solute removal. Such a spectacular, discordant relationship between rhythmic and cryptic layering was recently discovered in the layered Fongen gabbro, central Norway (Wilson & Larsen, 1982).

Fig. 39. In flat magma sheets, cooling rate at the extremities is high and allows little liquid diffusion. In the more interior portions of the magma, solute removal becomes increasingly important and thus implies changing phase assemblages which essentially follow the cooling rate isopleths. Solidification interfaces, on the other hand, are largely parallel to the contacts and thus possibly cross-cut the cooling rate isopleths, producing discordant relations between growth structures and phase assemblages.



6. Summary and conclusions

Directional solidification textures reveal important information about the conditions of crystallization which is not recognizable in isotropically textured products. Because the crystallization occurs sequentially from the contacts and inwards, analysis of thermal and compositional variations during directional solidification can be related to distance and time: systematic changes in crystal shapes are the result of different growth mechanisms which in turn reflect variable supercooling during crystallization; different mineral assemblages and their compositions reflect changing thermal conditions during the solidification; and finally, overall compositional variations permit evaluation of some kinetic properties of the constituent phases and the dynamic conditions in the source melt.

In the lardalite contact zones, crystal shapes near the margins are fan dendritic, the result of high thermal supercooling presumably following the emplacement of a hot magma into a cooler environment. Subsequent columnar growth with anisotropic, cellular textures imply growth by crystallographic control. At least one facet is required for the consistent orientation in the compact solid products and indicates that supercooling was above the roughening temperature of one crystallographic plane. Two facet planes subsequently promote the formation of columnar rod-shaped crystals, with growth restricted in two lateral directions. Unconstrained growth in the third dimension is directed parallel to heat flow. Consistent orientation of the 010 plane normal to the maximum growth direction in the columnar feldspars show that this face was the last to develop faceted growth, in accordance with its loose atomic packing and thus high roughening temperature. When growth is controlled by three

facet planes, the main elongation is controlled by crystallographic properties and deflected away from the direction of maximum heat flow; at this stage, the crystals start growing at an angle to the isotherms.

Growth in supercooled conditions results, in addition to low angle branching, in the formation of curved lattice planes which are readily recognised from curved polysynthetic twin planes and undulous extinction sweeping across individual subgrains upon rotation of the sample in plane polarized light. Apparently, the higher the supercooling during solidification is, the higher lattice curvature is achieved. The intersection of such curved lattice planes with growth facets, results in curved crystal shapes, formed in a completely static environment. Because of this curvature, the maximum growth is gradually directed away from the heat flow. When the growth of a curved dendrite becomes almost parallel to the isotherms, the isotherms move away from the solidification front, and repeated heterogenous nucleation occurs on top of the curved dendrite in order to readjust growth towards the heat flow. The dendritic crystallization thereby continuously adjust the solidification geometry to a maximum rate. Similarly, is the spacing of the dendrites continuously adjusted to optimum growth conditions. This happens because transverse diffusion between neighbouring dendrites readjusts the spacing to the actual growth conditions by eliminating excess dendrites, when the diffusion range is increasing, or allows subsequent bifurcation when the diffusion range is decreasing.

Directional growth occurs when a zone of constitutional supercooling is developed ahead of the solidification front. The changing crystal shapes in the lardalite contact zones show that thermal supercooling decreases continuously inwards whereas constitutional supercooling initially rises to a maximum and then decreases gradually until its ultimate extinction. When constitutional supercooling is eliminated, the crystallization produces equigranular textures. The presence of dendrites in some porphyritic textured zones, however, shows that the conditions for constitutional supercooling is eliminated sequentially for different phases, and thus that the porphyritic textures are formed by directional solidification despite the lack of textural evidence. This fact justifies the interpretation of equigranular and cumulus textured products in terms of lateral, directional solidification with repeated, sequential nucleation.

Compositional variations in the directionally solidified contact zones demonstrate that large compositional gradients are produced in a static boundary layer. The build up of solute and subsequent precipitation of complementary phases may lead to rhythmic layering comparable to

Skaergaard type layering. Also, remarkable examples of steady-state crystallization, which include compact, monomineralic layers, show that solute removal from the boundary layer occasionally balances the production at the solid-liquid interface. This type diffusion-controlled solidification occurs when the solid product approaches the composition of the liquid or when excess solute is removed by fluid flow, such as by convection near the crystallization front. The solute redistribution in these cases can be approximated by plane-front solidification models derived from binary systems by considering solvent as the bulk composition of the solid and solute as the sum of residual components.

In entirely equigranular products the direction of crystallization is not directly observable. The similarity of the contact zones with macrostructures of typical cast metal ingots, which include a chilled border zone, a columnar textured, marginal facies and an inner equigranular zone (Chalmers, 1964; Bolling, 1969; Flemings, 1974), suggests that the formation of equigranular textured pluton interiors is a product of directional solidification, and that the directional solidification of single phases in comb-textured contact zones is succeeded by repeated, heterogeneous nucleation of multifaceted crystals when constitutional supercooling is eliminated. In these conditions can compositional variations be ascribed to macrosegregation and dynamic solidification processes.

The concept of macrosegregation accounts for the overall compositional variation produced during solidification. Macrosegregation occurs primarily from fluid flow in the crystallization zone, and includes both laminar, convective flow along a planar solidification front and percolation flow in a mushy crystallization zone. This fluid flow serves to remove solute from the crystallization front. The flow interface, however, is invariably separated from the solid interface by a boundary layer of static liquid, in which transfer of solute occurs only by diffusion. The width of this static boundary layer controls the solute enrichment at the solid liquid interface, since the concentration of solute at the flow interface cannot exceed that of the flow unit; a wide boundary layer allows high solute concentration gradients whereas a narrow layer promotes minimum solute enrichment.

Crystallization with minimum supercooling in a magma at rest occurs mainly by heterogeneous nucleation along the cooler margins. Removal of heat of crystallization through solid media favours growth of crystals which are interconnected and attached to the wall rock relative to crystals in liquid suspension. The main solidification can therefore occur by a

chain nucleation and growth mechanism that creates a well defined solid-liquid interface, presumably subparallel to the isotherms.

In such a lateral solidification geometry, rejected solute accumulates ahead of the solidification front and is potentially removed by diffusion and fluid flow. Steady-state solidification may occur when solute enrichment in the boundary layer remains constant. This happens when a balance between solute build up at the solid interface and solute removal from the boundary layer at the flow interface is established. Quantitative modeling of the solute redistribution is possible when the solidification conditions are known. These conditions require knowledge of the diffusion coefficients, the solidification rate and width of the static boundary layer. Effective partition coefficients about unity readily form steady-state because the solute enrichment is minimum. High or low partition coefficients on the other hand require enhanced solute removal, such as by convective flow near the solid interface, for steady-state.

Extreme adcumulus growth can be treated as steady-state, monophasic precipitation with repeated nucleation; excess solute which would otherwise result in precipitation of secondary phases is continuously removed from the crystallization front by fluid flow along a narrow static boundary layer. If the removal of solute is limited, e.g. in a wide static boundary layer, periodic precipitation of secondary liquidus phases may lead to the formation of rhythmic compositional layering.

Macrosegregation is expressed by the compositional difference between the solid product and the source liquid and referred to as positive or negative. Initially large thermal supercooling near the contacts results in abundant nucleation and minimum segregation (chilled products). Later, when the cooling rate decreases, progressive removal of solute from the crystallization zone results in an increasingly negative segregation. Positive segregation (\sim differentiation) cannot precede negative segregation in a closed system. Magmas which show progressive, negative segregation towards the center (increasing adcumulus growth or reverse compositional zonation), accordingly must include a complementary liquid in order to account for the positive segregation in their petrogenesis.

Decreasing cooling rate towards the interior of a solidifying magma increases the time available for diffusion. A transition from predominantly orthocumulus growth (non-segregated) to predominantly adcumulus growth is therefore to be expected with continuous cooling. The result is a progressive, negative segregation. Because the solute

accumulation ultimately effects the source liquid, a final positive segregation can develop. In continuous castings with a planar solidification front, this positive segregation is usually experienced in restricted parts of the upper portion of the ingot, whereas the major, generally lower part suffers pronounced negative segregation.

Convection along a planar solidification front allows substantial amounts of the source magma to pass the crystallization zone and thus the precipitation of disproportionate amounts of selected phases, depending on their nucleation and growth characteristics. Because of the static boundary layer, the solute accumulation may attain any concentration level between that of the source liquid and the maximum enrichment at the solid interface; the solid phase assemblage thus 'sees' a source liquid whose composition depends only on the width of the boundary layer. Clearly, in such cases, the modelling of fractionation paths with the use of equilibrium phase relations has little significance for the evolution of the true source magma.

Lateral solidification with a planar or pseudoplanar front tends to follow the isotherms, which in turn reflect the cooling surface and the position of the solid-liquid interface. The formation of planar solidification structures such as rhythmic layering and textural lamination may therefore follow these planes and so define crystallization isochrons. Differentiation in the sense of macrosegregation, however, essentially reflects the dynamics of the solidification process, i.e. the cooling and fluid flow rates and the time available for solidification. Important differences in composition may therefore develop laterally if thermal gradients change along the solidification front; orthocumulus growth will predominate in steep thermal gradients whereas adcumulus growth controls more gentle crystallization conditions.

The discussion of compositional variations in magmas as a result of dynamic crystallization processes and macrosegregation is new to petrological thinking. In the previous decades main emphasis of petrogenetic modeling in plutonic rocks was placed on equilibrium phase relations, and differentiation based on various derivations of surface or total equilibrium crystallization models.

Fluid dynamical processes as a source of magma differentiation have been increasingly recognized in the literature, and are likely to contribute to the overall differentiation. Such processes include transport of crystals through the liquid by gravity induced convection (Irvine, 1979; Rice, 1981), double-diffusive convection (Turner, 1973, 1980; Turner & Gustavson, 1978; Chen & Turner, 1980; McBirney & Noyes, 1979; Huppert

& Turner, 1981), Soret or Dufour separation (Walker et al., 1981) and/or liquid immiscibility (Philpotts, 1976; Roedder, 1979). As the above processes essentially derive from liquid-liquid interactions they represent additional cases of magma evolution which are beyond the scope of this paper.

The solid-liquid interface, however, constitutes the most important transformation boundary which relates directly to the composition of the solid products. The present discussion is aimed at compositional and textural variations in the solid products which derive from processes at the solid-liquid interface, rather than on changes in the source liquids. This does not exclude important effects of the latter process, but the effects of such processes are considered secondary to those developed at the solid-liquid interface.

Rigorous quantitative approaches have deliberately been excluded from the text because, although quantitative solutions to many problems related to macrosegregation and crystallization kinetics are established for metallic systems, little experimental information is yet available for multicomponent silicate systems. The discussion, however, shows that a variety of important petrological phenomena, which include features that occasionally violate equilibrium phase relations, can be related to a single, dynamic solidification model.

ACKNOWLEDGEMENTS. I am grateful to Dr. Gary Lofgren, Experimental Petrology Branch at the Solar System Exploration Division, NASA, Johnson Space Center, Houston, for providing excellent research facilities during the tenure of a visiting fellowship. I wish to thank Dr. D. Gust, Lunar and Planetary Institute, and Dr. G. E. Lofgren, NASA-JSC for critically reading the manuscript and providing valuable suggestions. Dr. J. R. Wilson, Aarhus Univ., kindly commented upon early parts of the paper. Microprobe analyses at the Univ. of Oslo and NASA-SSED are also much appreciated.

References

- Allègre, C. J., Provost, A., & Jaupart, C., 1981: Oscillatory zoning: a pathological case of crystal growth. *Nature* 294, 223-228.
- Ashwal, L. D., 1982: Mineralogy of mafic and Fe-Ti oxide-rich differentiates of the Marcy anorthosite massif, Adirondacks, N.Y. *Am. Miner* 67, 14-27.
- Ashwal, L. D. & Seifert, K. E., 1980: Rare-earth-clement geochemistry of anorthosites and related rocks from the Adirondacks. *Bull. Geol. Soc. Am.* 91, 659-684.
- Baragar, W. R. A., 1960: Petrology of basaltic rocks in part of the Labrador Trough. *Bull. Geol. Soc. Am.* 71, 1589-1644.
- Bartlett, R. R., 1969: Magma convection, temperature distribution and differentiation. *Am. Jour. Sci.* 269, 1067-1982.
- Batemann, P. C. & Chappel, B. W., 1979: Crystallization, fractionation and solidification of the Tuolumne intrusive series, Yosemite National Park, Calif. *Bull. Geol. Soc. Am.* 90, 465-482.
- Bhattacharji, S., & Smith, C. H., 1964: Flowage differentiation. *Science* 145, 150-153.
- Bolling, G. F., 1969: Manipulation of structure and properties. In: *Solidification*. Amer. Soc. Metals, Metals Park Ohio, 341-385.
- Bowen, N. L., 1928: *Evolution of the Igneous Rocks*. Princeton Univ. Press.
- Brown, G. M., 1956: The layered ultrabasic rocks of Rhum, Inner Hebrides. *Phil. Trans. Roy. Soc. Lond. Ser-B*, 1-53.
- Bryan, W. B., 1972: Morphology of quench crystals in submarine basalts. *J. Geophys. Res.* 77, 5812-5219.
- Buerger, M. J., 1934: The lineage structure of crystals. *Z. Kristallogr. Miner.* 89, 195-220.
- Campbell, I. H., 1978: Some problems with the cumulus theory. *Lithos* 11, 311-323.
- Campbell, I. H., 1982: Layered intrusions: a mini review. In: D. Walker & I. S. McCallum (eds.) *Magmatic Processes of Early Planetary Crusts: Magma Oceans and Stratiform Layered Intrusions*. Lunar & Planet. Inst. T.Publ. 8201, 62-65.
- Carmichael, I. S. E., Turner, F. J., & Verhoogen, J., 1974: *Igneous Petrology*. McGrawhill.
- Chalmers, B., 1964: *Principles of Solidification*. J. Wiley & Sons.
- Chen, F. F. & Turner, J. S., 1980: Crystallization in a double-diffusive system. *J. Geophys. Res.* 85, 5B, 2573-2593.
- Demaiffe, D., & Hertogen, J., 1981: Rare earth element geochemistry and strontium isotope composition of a massif-type anorthositic-charnockitic body: The Hydra Massif. *Geochim. Cosmochim. Acta* 45, 1545-1561.
- Donaldson, G. H., 1976: An experimental investigation of olivine morphology. *Contr. Min. & Petrol.* 57, 187-213.
- Donaldson, C. H., 1977: Laboratory duplication of comb layering in the Rhum pluton. *Min. Mag.* 41, 323-336.
- Donaldson, C. H., 1983: Spinifex-textured komatiites: a review of textures, composition and layering. In: N. T. Arndt & E. G. Nisbet (eds.) *Komatiites*, 213-243.
- Donaldson, C. H., Usselman, T. M., William, R. J., & Lofgren, G. E., 1975: Experimental modeling of the cooling histories of Apollo 12 olivine basalts. *Proc. 6th Lunar Sci. Conf.*, 843-879.
- Donaldson, C. H., Drever, H. I., & Johnston, R., 1977: Supercooling on the lunar

- surface: a review of analogue information. *Phil. Trans. Roy. Soc. Lond. Ser.-A* 285, 207-217.
- Dowty, E., 1980: Crystal growth and nucleation theory and the numerical simulation of igneous crystallization. In: R. B. Hargraves (ed.) *Physics of Magmatic Processes*. Princeton Univ. Press, 419-486.
- Drever, H. J. & Johnston, R., 1972: Metastable growth patterns in some terrestrial and Lunar rocks. *Meteoritics* 7, 327-340.
- Elliott, R., 1977: Eutectic solidification. *Int. Metals. Rev.* 219, 161-186.
- Emmons, R. C. 1940: The contribution of differential pressures to magmatic differentiation. *Am. Jour. Sci.* 238, 1-21.
- Fenn, P. M., 1977: The nucleation and growth of alkali feldspars from hydrous melts. *Can. Miner.*, 15, 135-161.
- Flemings, M. C., 1974: *Solidification Processing*. McGraw-Hill.
- Gibb, F. G. F., 1974: Supercooling and the crystallization of plagioclase from a basaltic magma. *Min. Mag.* 39, 641-653.
- Gilmer, G. H., 1977: Computer simulation of crystal growth. *J. Cryst. Growth.* 42, 3-10.
- Grove, T. L., 1978: Cooling histories of Luna 24 very low Ti (VLT) ferrobasalts: an experimental study. *Proc. 9th Lunar Sci. Conf.* 565-584.
- Grove, T. L., & Walker, D., 1977: Cooling histories of Apollo 15 quartz normative basalts. *Proc. 8th Lunar Sci. Conf.* 1501-1520.
- Grove, T. L. & Beaty, D. W., 1980: Classification, experimental petrology and possible volcanic histories of the Apollo 11 high-K basalts. *Proc. 11th Lunar Sci. Conf.* 149-177.
- Harker, A., 1909: *The Natural History of Igneous Rocks*. Macmillan.
- Hartman, P., 1973: Structure and morphology. In: P. Hartman (ed) *Crystal Growth: An Introduction*. North Holland, 367-402.
- Hartman, P., 1982: Crystal faces: structure and growth. *Geol. Mijnbouw*, 61, 313-320.
- Henderson, P., 1970: The significance of the mesostasis of basic layered igneous rocks. *J. Petrology* 11, 463-473.
- Henderson, P., & William, C. T., 1977: Variations in trace element partitioning as function of crystal growth rate. In: L. H. Ahrens (ed.) *Origin and Distribution of the Elements 2. Symp.*, 191-198.
- Hess, H. H., 1938: Primary banding in norite and gabbro. *Trans. Am. Geophys. Union 19th Ann. Mtg.*, 264-268.
- Hildreth, W., 1979: The Bishop Tuff: Evidence for the origin of compositional zonation in silicic magma chambers. *Geol. Soc. Am. Sp. Pap.* 180, 43-75.
- Hofmann, A. W., 1980: Diffusion in natural silicate melts: a critical review. In: R. B. Hargraves (ed.) *Physics of Magmatic Processes*, Princeton Univ. Press, 385-418.
- Howarth, J. A., & Mondolfo, L. F., 1962: Dendritic growth. *Acta Metall.* 10, 1037-1042.
- Hunt, J. D., & Hurlle, T. J., 1968: The structures of faceted/nonfaceted eutectics. *Trans. Metall. Soc. AIME* 242, 1043-1047.
- Huppert, H. E., & Turner, J. S., 1981: Double diffusive convection. *J. Fluid Mech.* 106, 299-329.
- Irvine, T. N., 1979: Rocks whose composition is determined by crystal accumulation and sorting. In: H. S. Yoder (ed.) *The Evolution of the Igneous Rocks. Fiftieth Anniversary Perspectives*. Princeton Univ. Press, 245-306.

- Irvine, T. N., 1980: Magmatic infiltration metasomatism, double-diffusive fractional crystallization and adcumulus growth in the Muskox intrusion and other layered intrusions. In: R. B. Hargraves (ed.) *Physics of Magmatic Processes*, Princeton Univ. Press, 325-384.
- Jackson, E. D., 1961: Primary textures and mineral associations in the ultramafic zone of the Stillwater Complex, Montana. *Prof. Pap. U.S. Geol. Surv.* 358.
- Jackson, K. A., 1958: Mechanism of growth. In: R. H. Doremus, B. W. Roberts & P. Turnbull (eds.) *Growth and Perfection of Crystals*. J. Wiley & Sons, 319-359.
- Jackson, K. A., 1972: Defect formation, microsegregation and crystallization morphology. In: *Solidification*. Amer. Soc. Metals., Metals Park Ohio, 121-154.
- Jackson, K. A. & Hunt, J. D., 1966: Lamellar and rod eutectic growth. *Trans. Metall. Soc. AIME* 236, 1129-1142.
- Jaeger, J. C., 1968: Cooling and solidification of igneous rocks. In: H. H. Hess (ed.) *Basalts. Vol. 2*; Wiley, 503-536.
- Jahns, R. H. & Tuttle, O. F., 1963: Layered pegmatite-aplite intrusives. *Min. Soc. Am. Spec. Pap. 1*, 78-92.
- Keith, H. D. & Padden, F. J., 1963: A phenomenological theory of spherulitic crystallization. *J. Appl. Phys.* 34, 2409-2421.
- Kirkpatrick, R. J., 1975: Crystal growth from the melt: a review. *Am. Miner.* 60, 798-814.
- Kirkpatrick, R. J., 1981: Kinetics of crystallization of igneous rocks. In: A. C. Lasaga & R. J. Kirkpatrick (eds.) Kinetics of geochemical processes. *Min. Soc. Am. Rev. in Min.* 8, 321-398.
- Kirkpatrick, R. J., 1983: Theory of nucleation in silicate melts. *Am. Miner.* 68, 66-77.
- Komar, P. D., 1972: Mechanical interaction of phenocrysts and flow differentiation of igneous dikes and sills. *Bull. Geol. Soc. Am.* 83, 973-988.
- Lofgren, G. E., 1974a: An experimental study of plagioclase crystal morphology: isothermal crystallization. *Am. Jour. Sci.* 274, 243-273.
- Lofgren, G. E., 1974b: Temperature induced crystallization in synthetic plagioclase feldspars. In: W. S. Mackenzie & J. Zussmann (eds.): *The Feldspars*. Manchester Univ. Press, 362-375.
- Lofgren, G. E., 1977: Dynamic crystallization experiments bearing on the origin of textures in impact generated liquids. *Proc. 8th Lunar Sci. Conf.* 2079-2095.
- Lofgren, G. E., 1980: Experimental studies on the dynamic crystallization of silicate melts. In: R. B. Hargraves (ed.) *Physics of Magmatic Processes*. Princeton Univ. Press, 487-552.
- Lofgren, G. E., 1983: Effect of heterogeneous nucleation on basaltic textures: a dynamic crystallization study. *J. Petrology* 24, 229-255.
- Lofgren, G. E., Donaldson, G. H., Williams, R. J., Mullins, O. & Usselman, T. M., 1974: Experimentally reproduced textures and mineral chemistry of Apollo 15 Qz-normative basalts. *Proc. 5th Lunar Sci. Conf.*, 549-567.
- Lofgren, G. E. & Donaldson, G. H., 1975: Curved, branching crystals and differentiation in comb-layered rocks. *Contr. Miner. Petrol.* 49, 309-319.
- Lofgren, G. E. & Gooley, R., 1977: Simultaneous crystallization of feldspar intergrowth from the melt. *Am. Miner.* 62, 217-228.
- McBirney, A. R., & Noyes, R. M., 1979: Crystallization and layering of the Skaergaard Intrusion. *J. Petrology*, 20, 487-554.

- McIntire, W. L., 1963: Trace element partition coefficients – a review of theory and application to geology. *Geochim. Cosmochim. Acta* 27, 1209-1264.
- Mehrabian, R., 1984: A review of our present understanding of macrosegregation in ingots (abs.) In: *Fundamentals of Alloy Solidification Applied to Industrial Processes*. NASA Conf. Publ. 2337, 169-185.
- Michot, J & Michot, P., 1969: The problem of anorthosites: The south Rogaland igneous complex, SW Norway. In: Y. W. Isachsen (ed.) *Origin of Anorthosites and Related Rocks*. N. Y. State Mus. Serv. Mem. 18, 399-410.
- Miller, C. E., 1977: Faceting transition in melt-grown crystals. *J. Cryst. Growth* 42, 357-363.
- Mollard, F. R., & Flemings, M. C., 1967: Growth of composites from the melt – part 1. *Trans. Metall. Soc. AIME* 239, 1526-1533.
- Moore, J. G., & Lockwood, J. P., 1973: Origin of comb layering and orbicular structure. Sierra Nevada Batholith, California *Bull. Geol. Soc. Am.* 84, 1-20.
- Morris, L. R., & Winegaard, W. C., 1969: The cell to dendrite transition. *J. Cryst. Growth* 6, 61-66.
- Morse, S. A., 1970: Alkali feldspars with water at 5 kbar pressure, *J. Petrology* 11, 211-253.
- Morse, S. A., 1979: Influence of augite on plagioclase fractionation. *J. Geol.* 87, 202-208.
- Morse, S. A., 1980: *Basalts and Phase diagrams*. Springer-Verlag.
- Nesbitt, R. W., 1971: Skeletal crystal forms in the ultramafic rocks of the Yilgarn Block, Western Australia: evidence for an Archaean ultramafic liquid. *Geol. Soc. Aust. Sp. Publ.* 3, 331-348.
- Neumann, E.-R., 1976: Compositional relations among pyroxenes, amphiboles and other mafic phases in the Oslo region plutonic rocks. *Lithos* 9, 85-109.
- Oftedahl, C., 1948: Studies on the igneous rocks complex of the Oslo region. IX. The feldspars. *Skr. Norske. Vid.-Akad. Oslo I. Mat.-naturv. Kl.* 3.
- Petersen, J. S., 1978a: Structure of the larvikite-lardalite complex, Oslo region, Norway, and its evolution. *Geol. Rundsch.* 67, 330-342.
- Petersen, J. S., 1978b: Southern part of the Oslo-rift. In: J. A. Dons & B. T. Larsen (eds.) *The Oslo Paleorift. Nor. Geol. Unders.* 337, 171-182.
- Petersen, J. S., 1985: Columnar-dendritic feldspars in the lardalite intrusion, Oslo region, Norway: 1 Implications for unilateral solidification of a stagnant boundary layer. *J. Petrology* 26, 223-252.
- Philpotts, A. R., 1976: Silicate liquid immiscibility: Its probable extent and petrogenetic significance. *Am. Jour. Sci.* 276, 1147-1177.
- Platten, I. M., & Watterson, J. S., 1969: Oriented crystal growth in some tertiary dykes. *Nature* 223, 286-287.
- Poldervaart, A. & Taubeneck, W. H., 1959: Willow-Lake type layered intrusions. *Bull. Geol. Soc. Am.* 70, 1393-1398.
- Rice, A., 1981: Convective fractionation: A mechanism to provide cryptic zoning (macrosegregation), layering, crescumulates, banded tuffs and explosive volcanism in igneous processes. *Jour. Geophys. Res.* 86; B1, 405-417.
- Rinne, F., 1926: Thermotaxie als problem der orientierten kristallisation. *Zeitsch. Krist.* 64, 71-75.
- Roedder, E., 1979: Silicate liquid immiscibility in magmas. In: H. S. Yoder (ed.) *The Evolution of the Igneous Rocks. Fiftieth anniversary prespective*. Princeton Univ. Press, 15-57.

- Rohatgi, P. K. & Adams, C. M. Jr., 1967: Effect of freezing rates on dendritic solidification of ice from aqueous solutions. *Trans. Metall. Soc. AIME* 239, 1729-1736.
- Rutter, J. W. & Chalmers, B., 1953: A prismatic substructure formed during solidification of metals. *Can. J. Phys.* 31, 15-39.
- Scherer, G. W. & Uhlmann, D. R., 1975: Diffusion-controlled growth of dendrite arrays. *J. Cryst. Growth* 30, 304-310.
- Schiffmann, P. & Lofgren, G. E., 1982: Dynamic crystallization studies of the Grande Ronde pillow basalts, central Washington. *J. Geol.* 90, 49-78.
- Shannon, J. R. Walker, B. M. Carter, R. B. & Geraghty, E. P., 1982: Unidirectional solidification textures and their significance in determining relative ages of intrusion at the Henderson Mine, Colorado. *Geology* 10, 293-297.
- Shewmon, P. G. 1969: *Transformations in Metals*. McGraw-Hill.
- Swanson, S. E., 1978: Petrology of the Rocklin pluton and associated rocks, western Sierra Nevada, Calif. *Bull. Geol. Soc. Am.* 89, 679-686.
- Taubeneck, H. & Poldervaart, A., 1960: Geology of the Elkhorn Mts., Northeast Oregon: 2 Willow Lake intrusion. *Bull. Geol. Soc. Am.* 71, 1295-1322.
- Tiller, W. A. Jackson, K. A. Rutter, J. W., & Chalmers, B., 1953: The redistribution of solute atoms during the solidification of metals. *Acta Metall.* 1, 428-437.
- Turner, J. S., 1973: *Buoyancy Effects in Fluids*. Cambridge Univ. Press.
- Turner, J. S., 1980: A fluid-dynamical model of differentiation and layering in magma chambers. *Nature* 285, 213-215.
- Turner, J. S. & Gustavson, L. B., 1978: The flow of hot solutions from vents in the sea floor – some implications for exhalative massive sulfide and other ore deposits. *Econ. Geol.*, 73, 1082-1100.
- Usselman, T. M., Lofgren, G. E., Donaldson, C. H. & Williams, R. H., 1975: Experimentally reproduced textures and mineral chemistries of high-titanium mare basalts. *Proc. 6th Lunar Sci. Conf.*, 997-1020.
- Voensdreght, C. F., 1983: Crystal morphology of monoclinic potassium feldspars: a qualitative approach with special emphases on the periodic bond chain theory of Hartman & Perdok. *Z. Kristall.*
- Wadsworth, W. J., 1960: The ultrabasic rocks of southwest Rhum. *Phil. Trans. Roy. Soc. Lond. Ser.-B* 244, 21-64.
- Wager, L. R., Brown, G. M. & Wadsworth, W. J., 1960: Types of igneous cumulates. *J. Petrology* 1, 73-85.
- Wager, L. R. & Brown, G. M. 1967: *Layered Igneous Rocks*. Oliver & Boyd.
- Wager, L. R. & Deer, W. A., 1939 (re-issued 1962): Geological Investigations in East Greenland, III. The Petrology of the Skaergaard Intrusion, Kangerdlugssuaq, East Greenland. *Medd. om Grønland* 105, No. 4, 1-352.
- Walker, D., Kirkpatrick, R. J., Longhi, J., & Hayes, J. F., 1976: Crystallization history of lunar picrite basalt sample 12002: phase equilibria and cooling-rate studies. *Bull. Geol. Soc. Am.* 87, 646-656.
- Walker, D., Leshner, C. E., & Hayes, J. F. 1981: Soret separation of lunar liquid. *Proc. Lunar Planet. Sci. Conf. XIIB*, 991-999.
- Wilson, J. R & Larsen, S. B., 1982: Discordant layering in the Fongen-Hyllingen basic intrusion. *Nature* 299, 625-626.
- Wright, T. L. & Doherty, P. C., 1970: A linear programming and least squares computer method for solving petrological mixing problems. *Bull. Geol. Soc. Am.* 81, 1995-2008.

JENS KONNERUP-MADSEN
*Composition of Gases
in the Earth's Upper Mantle*

ABSTRACT. There is considerable evidence suggesting the continuous release of gases from deeper parts of the Earth. A review is given of geological data pertinent to deep gas compositions, emphasizing information of relevance to upper mantle conditions, and to possible degassing mechanisms. Data from experimental and theoretical studies relating to the existence and composition of upper mantle gases are presented, and information on deep gas compositions from natural rocks, comprising volcanic gases, gases in ocean-floor basalt glasses, in phenocrysts and upper mantle minerals, is reviewed. The available evidence is not decisively in favour of any single upper mantle gas composition. H₂O and CO₂ appear to be the major species, whereas more reduced gases are present in only subordinate concentrations and are of only minor importance for Earth degassing processes. Very low total gas contents, in the order of less than 0.1 weight % gases for some upper mantle regions, are indicated, but an upper mantle heterogeneous in its gas content and composition seems likely.

Institute of Petrology, University of Copenhagen,
Østervoldgade 10, DK-1350 Copenhagen K, Denmark

Introduction

Approximately 20% of the total volume of unmetamorphosed sediments at the Earth's surface consists of pore water. During subsequent burial of such sediments most of this water (and other gases, e.g. formed from reactions with organic materials present in the sediments) are expelled from the sediments and progressively lost from the sedimentary pile during diagenesis and metamorphism. Therefore most metamorphic rocks at the base of the Earth's crust are low in gases such as CO₂ and H₂O, with the gases bound in minerals such as amphiboles, micas, and carbonates. Partly because of this trend toward lower contents of gases with depth it is less commonly recognized that most geological processes deeper in the Earth's crust and upper mantle have taken place in the presence of, or have been influenced by, gases of various compositions.

Studies of natural rocks document that gases, either bound in minerals

or as a free gas phase, are actually present in the deeper parts of the Earth. Much of this evidence comes from volcanic activity. Water (H_2O) and carbon dioxide (CO_2) are frequently reported as important constituents of the gases released during volcanic eruptions and are therefore often judged also to be the most important gases in the lower crust and upper mantle. Recently, however, hypotheses of the existence of more reduced gases such as methane (CH_4) in the deeper parts of the Earth's crust and upper mantle have been revived (Gold, 1979; MacDonald, 1983) and have questioned the accuracy of the above simple picture of deep Earth gas compositions.

At a theoretical and experimental level the importance of gases for rock forming processes in general is well documented. In the Earth's crust the presence of gases has important implications for mass transport, deformation mechanisms, rheological properties, and for mineral reactions and equilibria. At still deeper levels, in the upper mantle, gases may significantly affect mineral phase equilibria important for the formation and subsequent evolution of silicate magmas. The presence of deep-seated gases and their possible migration to the Earth's surface may furthermore have important implications for the chemical evolution of the Earth's atmosphere and hydrosphere (Cogley and Henderson-Sellers, 1984).

It is the aim of this article to review the available geological evidence for the presence and compositions of gases, and their actual concentrations and distribution in the Earth's upper mantle. Emphasis has been placed as much on the rationale behind the pertinent geological studies as on the obtained results themselves. The greatest importance will be attached to results obtained from studies of gases in natural volcanic rocks, and to degassing phenomena associated with their formation. As a basis for these data a short account of experimental and theoretical constraints on compositions of deep gases will be given.

The Earth's crust and upper mantle

Some major features of the Earth's crust and upper mantle of relevance to this article are shown schematically in Fig. 1.

The crust consists of the region above the Mohorovicic discontinuity which is defined by an abrupt increase in seismic P-wave velocities considered to reflect a change in density of the rocks due to a mineralogical or chemical change or both. The crust generally ranges in thickness from

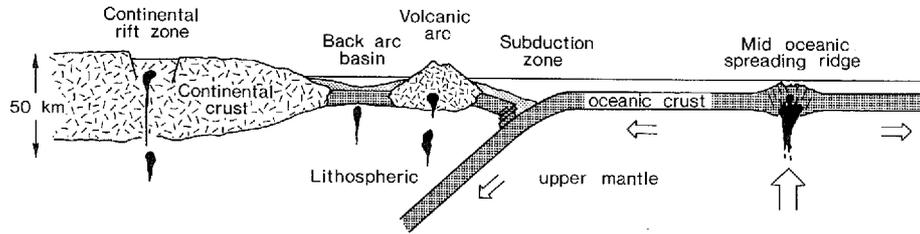


Fig. 1. Schematic cross-section of the Earth's crust and uppermost mantle showing the main large-scale geological features referred to in this article and placed in a plate-tectonic context. Arrows indicate directions of relative movement of the oceanic plate. Mantle derived magmas are shown by black, sediments are light stippled.

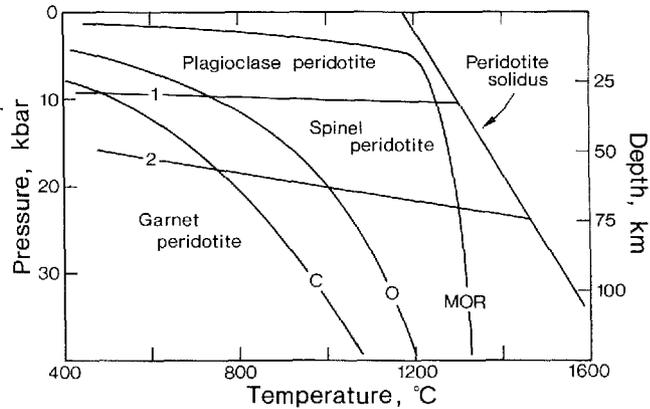
about 5-15 km in oceanic regions to 30-50 km in continental regions. Oceanic islands, island arcs and continental margins are examples of transitional crust that exhibit thicknesses of 15-20 km.

The upper mantle extends from the Mohorovicic discontinuity to a depth of about 400 km and comprises the lithosphere and the upper part of the asthenosphere. The lithosphere is between 50 and 150 km thick and forms an upper more brittle part of the upper mantle. The asthenosphere extends from the base of the lithosphere and down to about 700 km. The uppermost 50-150 km of the asthenosphere is characterized by relatively low seismic-wave velocities and forms the low-velocity zone. The asthenosphere deforms plastically. These differences in physical properties of the lithosphere and asthenosphere are reflected in the plate tectonic model, where the Earth's surface is considered to be made up of a number of rigid plates which are in motion relative to each other (e.g. Wyllie, 1971). The crust and the lithosphere together form the rigid plates moving on top of the more plastic asthenosphere, where the convection cells moving the plates are believed to be concentrated.

New oceanic crust is constantly being formed along mid oceanic rift zones where basaltic melts generated by the upward movement and subsequent partial melting of deep mantle materials are erupted. Generation of magma in the mantle also occurs where older oceanic crust is subducted under continental or oceanic crust, with ascent of the derivative magmas close to the subduction zone and the formation of volcanic island arcs. Continental analogues of island arcs occur in regions (e.g. the Andes in South America) where subduction of oceanic crust occurs beneath a continental mass.

The continental crust is dominated by a variety of different metamor-

Fig. 2. Vertical zonation of a gas-free peridotitic mantle, as inferred from experiments. The beginning of melting of solid mantle is indicated by the peridotite solidus. Curves 1 and 2 indicate the positions of mineral reactions separating depth-regions for spinel-bearing peridotite from plagioclase- and garnet-bearing peridotite (after Wyllie, 1971). Curves C, O and MOR show geotherms for continental (C), oceanic (O) and mid oceanic ridge (MOR) regions, respectively (from Anderson, 1981).



phic and plutonic rock types and a cover of sedimentary rocks of variable thickness. Within some stable continental blocks there are rift systems, which are fault-bounded valleys ranging in width from 30–70 km and in length from a few tens to thousands of kilometers. Such continental rift systems often have associated volcanic activity, with extrusion of lavas originating deep within the upper mantle. Continental rift zones mark the splitting of continental plates and may develop into new oceanic crust.

Whereas samples of the Earth's crust are easily accessible for collection and study, samples of the upper mantle are more fragmentary in nature and essentially consist of the melted portion of the mantle erupted as lavas and the solid fragments of presumed mantle rocks (xenoliths) they occasionally take with them. Although studies of available mantle samples have established the heterogeneous character of the upper mantle (Dawson, 1981) there is a general agreement that the rock type peridotite forms a substantial part of the upper mantle, the major mineralogy being olivine, two pyroxenes (ortho- and clinopyroxene), plus an aluminous phase changing from plagioclase to spinel and then to garnet with increasing depth (pressure), as visualized in Fig. 2.

Experimental and theoretical constraints on deep gas compositions

The gas-free peridotitic upper mantle of Fig. 2 is contradicted by several observations on natural rocks. The occurrence of amphiboles and mica in upper mantle xenoliths brought to the Earth's surface by kimberlitic magmas and alkaline lavas demonstrates that water at least locally is present in the mantle (Dawson and Smith, 1982). Similarly, carbonates as inclusions in garnet and olivine in deep mantle samples indicate the presence of CO_2 (Dawson, 1980). Additional evidence for the potential importance of gases other than CO_2 and H_2O is given by analyses of the volatiles released during volcanic eruptions (Nordlie, 1971). H_2O , CO_2 , SO_2 and H_2 are the main volcanic gas components and although such gases may not represent the actual gas speciation in a deep gas phase the importance of some species in the system C-O-H-S is clearly indicated (Fig. 3).

Partly because H_2O and CO_2 after all are considered to be the most important geological gases, partly because of experimental difficulties in controlling exactly the gas species composition in multispecies fluids at high temperatures and pressures practically all experiments have focused on the influence of H_2O and CO_2 on mantle phase relations. In keeping with these conditions most of the data presented in subsequent sections deal with gas species within the subsystem C-O-H (front of Fig. 3).

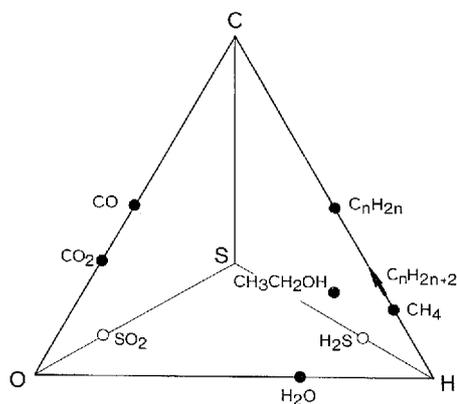


Fig. 3. Main gas species reported present in minerals and rocks of upper mantle derivation plotted in the C-O-H-S system.

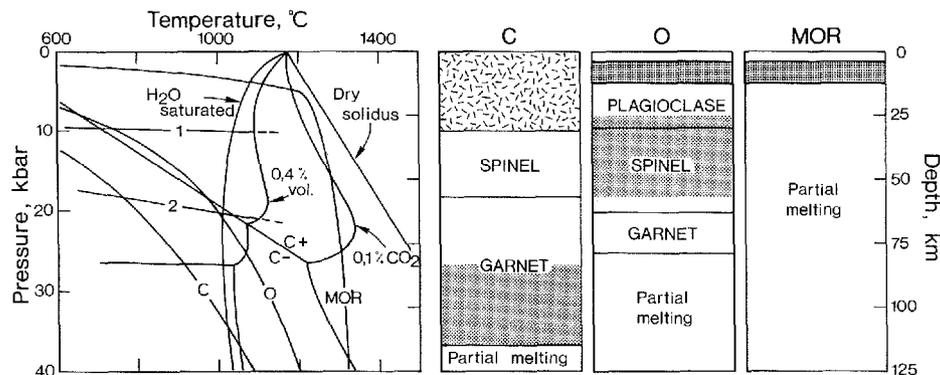
A peridotite-CO₂-H₂O upper mantle

The generalized phase relations for peridotitic mantle containing CO₂ and H₂O are given in Fig. 4.

Addition of H₂O and/or CO₂ to the dry system causes temperatures for beginning of melting of peridotitic mantle to decrease from those when no gas is present (Mysen and Boettcher, 1975; Wyllie, 1979). The beginning of melting curves (solidi) for three conditions of gas concentrations are sketched in Fig. 4 (left diagram). Under sub-solidus conditions addition of H₂O and CO₂ allows the formation of hydroxyl bearing minerals and carbonates, respectively. From Fig. 4 (left diagram) the mineral assemblages present at various depths can furthermore be determined by following the three geothermal gradients, representing the main geological settings, through the various mineral reaction and melting curves of Fig. 4. Mantle cross sections so derived are shown in the three columns to the right in Fig. 4 for an intermediate gas content of 0.3 weight % H₂O plus 0.1 weight % CO₂ and *a priori* considered to reflect plausible mantle gas concentrations (Olafsson, 1980).

Fig. 4. Vertical zonation of a peridotite-CO₂-H₂O upper mantle. The left pressure-temperature diagram shows solidus curves for peridotite-CO₂-H₂O for the volatile contents stipulated in the diagram and compared with geotherms from Fig. 2. The curve C-/C+ (= carbonates/free CO₂) indicates a sub-solidus carbonation reaction (Wyllie and Huang, 1975; 1976). Other abbreviations as in Fig. 2.

The three right columns show vertical cross-sections for a peridotite-CO₂-H₂O mantle with 0.4 weight % volatiles (0.3 wt.% H₂O + 0.1 wt.% CO₂) for continental (C), oceanic (O) and mid oceanic ridge (MOR) regions. Mineralogical zones, and depth ranges for existence of a free gas phase (light stippled) and partial melting of mantle are indicated. Uppermost in the sections are indicated the average thicknesses of continental (irregular strokes) and oceanic (heavy stippled) crust, and ocean. Diagrams based on data in Mysen and Boettcher (1975) and Olafsson (1980).



An important result indicated by the three mantle cross sections (Fig. 4) is that upon increasing the geothermal gradient (from continental towards mid oceanic ridge regions) a decrease in both the depth to partial melting of mantle and to regions with the possible existence of a free gas phase takes place. Thus, beneath stable continental crustal regions a free gas phase may exist at depths between 80 and 130 km, due to the breakdown of the OH-bearing phase phlogopite, and resulting in a H₂O-rich gas phase. The CO₂ under continental areas is stored in carbonate. In oceanic regions, because of the higher geothermal gradient, melting occurs at shallower depths and a free gas phase may exist within the depth range 25-55 km and have a more intermediate CO₂-H₂O composition (Wyllie, 1979). Under mid-ocean ridge regions a free gas phase, very rich in CO₂, may exist close to the surface and partial melting occurs at all depths greater than about 20 km.

The carbonation reaction (C-/C+) in Fig. 4 is actually only one of three sub-solidus divariant carbonation reactions identified in the model system CaO-MgO-SiO₂-CO₂ by Wyllie and Huang (1975; 1976). According to these model reactions it takes about 5 weight % CO₂ to react away all the clinopyroxene in a normal mantle peridotite, and about 23 weight % CO₂ to transform a peridotite completely into coesite-garnet marble. In some garnet kimberlites (Dawson, 1980) all clinopyroxene and most of the olivine appears to have reacted away and thus indicates that high CO₂ concentrations, at least locally, can occur in the upper mantle.

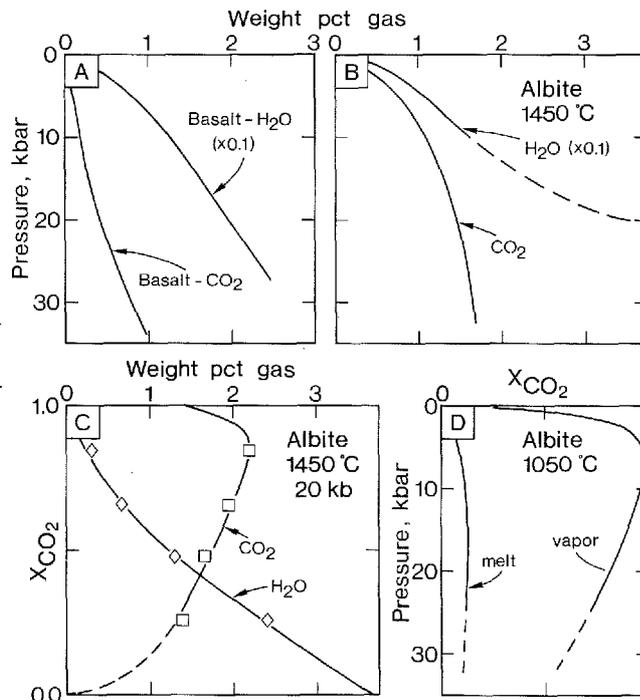
Amphibole is formed in peridotite by reaction with water. Up to about 0.4 weight % H₂O can be stored in amphibole-peridotite, and up to about 5 weight % CO₂ in dolomite-peridotite without losing one or more of the principal peridotite minerals.

The results obtained from Fig. 4 furthermore rest on the assumption that any gas species is only stoichiometrically incorporated into minerals such as amphiboles, micas and carbonates. Studies have, however, indicated that species within the system C-O-H may be incorporated at levels of tens to hundreds of parts per million into stoichiometrically volatile-free minerals such as olivine, pyroxene and garnet (e.g. Freund et al., 1980; Aines and Rossman, 1984), and that amphiboles from mantle xenoliths may contain up to about 0.1-0.5 weight % carbon-bearing gas species in addition to water (Matson and Muenow, 1984). Therefore a free gas phase may not exist even though it would be predicted from experimental studies.

Fig. 5. Solubility relations for CO_2 and H_2O in basalt and albite melts. (A) Solubility of CO_2 and H_2O in basalt melt along the basalt- CO_2 and basalt- H_2O solidus curves, respectively, as a function of pressure (/depth). Data from Wyllie (1979) and Spera and Bergman (1980). (B) Solubility of CO_2 and H_2O in albite melt at 1450°C as a function of pressure. Data from Burnham and Davis (1974) and Mysen et al. (1976)

(C) Solubility of CO_2 and H_2O in $(\text{CO}_2 + \text{H}_2\text{O})$ -saturated albite-melt at 1450°C and 20 kbar as a function of the mole fraction of CO_2 (X_{CO_2}) in the system (from Mysen, 1976).

(D) Compositions, in terms of mole fraction of CO_2 (X_{CO_2}), of coexisting vapour and melt in equilibrium with albite at 1050°C as function of pressure at CO_2 and H_2O saturation of the melt (from Eggler and Kadik, 1979).



Solubilities of gases in upper mantle melts and minerals

The most important features concerning the solubilities of H_2O and CO_2 in melts at temperatures and pressures relevant to upper mantle conditions are shown in Fig. 5.

On the basis of the results presented in Fig. 5 the following general comments regarding the role of CO_2 and H_2O for melting of upper mantle rocks can be given:

(1) At the same temperature, pressure and melt composition 3-10 times as much H_2O as CO_2 can dissolve in the melt. At pressures of about 20 kbar, corresponding to a depth of 65 km, about 20 weight per cent H_2O and only about 0.5 weight CO_2 can dissolve in tholeiitic basalt melt (Fig. 5,A). Higher solubilities are found for alkali basalt melt compositions (Mysen et al., 1976).

(2) The solubilities of both CO_2 and H_2O are positively correlated with pressure (Fig. 5,B) whereas the correlation with temperature ap-

pears to be linear or at lower pressures possibly negative (Rai et al., 1983).

(3) The presence of H_2O affects CO_2 solubilities; about 20-30% more CO_2 dissolves in hydrous than in H_2O -free silicate melts (Fig. 5,C).

(4) In the presence of a mixed ($\text{CO}_2 + \text{H}_2\text{O}$)-gas phase, because of the very different solubilities of CO_2 and H_2O in any melt composition, the gas phase is strongly enriched in CO_2 over gases dissolved in the coexisting melt (Fig. 5,D).

A consequence of this fractionation of gases (Fig. 5,D) is that partial melting of a ($\text{CO}_2 + \text{H}_2\text{O}$)-bearing peridotite may result in enrichment of the residual mantle in CO_2 whereas the melt will be enriched in H_2O . Thus partial melting may result in preferential depletion in H_2O in the magma-forming regions of the upper mantle and thereby create a mantle that is heterogeneous with respect to gas components. Furthermore, although the gas composition of the derivative melt will be highly enriched in H_2O any gas exsolved at lower pressures from the melt because of gas saturation will be highly enriched in CO_2 and differ considerably in composition from the initially dissolved gas phase. Using albite as a model (Fig. 5,D) almost pure CO_2 would exsolve from a ($\text{CO}_2 + \text{H}_2\text{O}$) saturated melt in the pressure range 5-10 kbar.

Considering the high CO_2 and H_2O concentrations needed to saturate a melt at upper mantle conditions, it appears realistic to conclude that most, if not all, melts formed in the upper mantle are unsaturated in H_2O and CO_2 , with complete and instantaneous absorption of any free gas phase by the magma in the magma-forming regions (Kadik and Lukin, 1973).

Studies of the solubility of CO_2 /carbon in minerals coexisting with a CO_2 gas at mantle conditions are few and the results inconclusive. Mysen et al. (1976) found solubilities of carbon in olivine and diopside at 20-30 kbar to be 7-10 ppm. Such low carbon solubilities would preclude storage of any significant amounts of carbon (dioxide) in the upper mantle within crystalline silicates and suggest that any mantle CO_2 is stored in a gas phase, in carbonates, in the form of graphite/diamond, or in silicate melts. However, much higher carbon contents, in the order of 500-2000 ppm, have been obtained by Mathez (1984) and Freund et al. (1980) for mantle olivine, and carbon contents of 140-400 ppm have been obtained on kaersutites from mantle xenoliths (Poreda and Basu, 1984).

Experimental data on solubilities of gases other than CO_2 and H_2O are scarce and at best available only for conditions not representative for the upper mantle. Egger et al. (1979) measured the solubility of CO_2 -CO

volatile mixtures in equilibrium with graphite in various melt compositions at pressures of 20-30 kbar and 1700 °C and observed slightly higher solubilities than for CO₂ alone. Extrapolating available data on other gases to a pressure of 2 kbar enables comparison with the solubilities of CO₂ and H₂O, and the following solubilities (in weight %) were obtained: He (0.0043%), Ar (0.0071%), Ne (0.0126%), N₂ (0.037%), CO₂ (0.55%), and H₂O (4.4%), with data from Kirsten (1968) and Kesson and Holloway (1974).

Sulfur behaves slightly different than CO₂ and H₂O in that sulfur solubility decreases with increasing pressure and increases with increasing temperature and total FeO-content of a silicate melt (Wendlandt, 1982). In the range 20-30 kbar sulfur solubilities are in the order of 0.1-0.25 weight %.

Gas species composition of a free C-O-H upper mantle gas phase

Evidence for the oxidation state of the upper mantle is not decisive in favor of any single value. Information has essentially been obtained from, either thermodynamic and experimental studies relating to chemical compositions of coexisting minerals in upper mantle xenoliths (e.g. Eggler, 1983), or direct intrinsic oxygen fugacity measurements on mantle minerals (e.g. Arculus and Delano, 1980). Whereas results from the latter type of approach suggest a strongly reduced upper mantle (with oxygen fugacities corresponding to the iron – wüstite buffer), results from the former method of study indicate slightly more oxidized conditions, with oxygen fugacities corresponding to the quartz – fayalite – magnetite buffer assemblage. The actual oxidation state of the upper mantle is important as the species composition of a free gas phase will be very dependent on the actual oxygen fugacity level.

In Fig. 6 the influence of the oxidation state for the compositions of a free C-O-H gas in equilibrium with graphite/diamond is given for pressure-temperature conditions corresponding to the upper mantle in the Lesotho kimberlite province (Fig. 6,A). Oxygen fugacities calculated from the chemical compositions of coexisting olivine, pyroxene and oxides in high-pressure xenoliths from various kimberlites and peridotites from the upper mantle are shown in Fig. 6,B (Eggler, 1983). Values cluster between the quartz – fayalite – magnetite and magnetite – wüstite oxygen buffer curves and are nearly centred on the oxygen fugacity values defined by the coexistence of enstatite – magnesite – olivine – graphite/diamond (Eggler et al., 1980). These relatively oxidized conditions may be interpreted as representing the oxidation state of the sub-

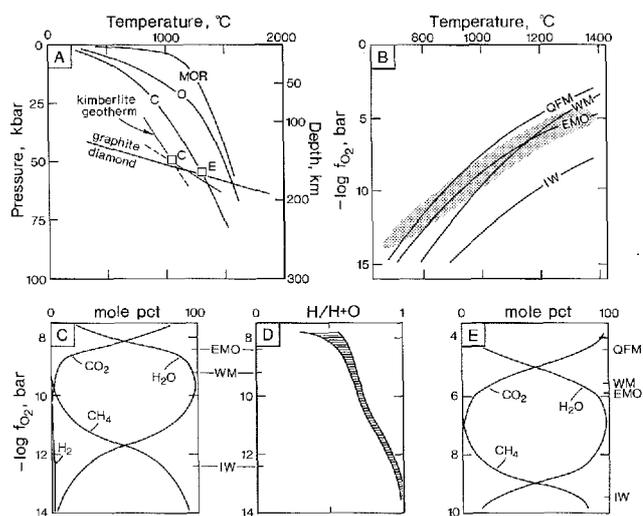


Fig. 6. Gas species compositions of a free C-O-H gas phase as function of temperature, pressure and level of oxidation of the upper mantle. (A) Pressure-temperature diagram showing geotherms (C, O and MOR) from Fig. 2 and the inflected sub-continental geotherm proposed for the Lesotho kimberlite province (Boyd, 1973). Graphite-diamond stability curve from Bundy et al. (1961). Boxes C and E indicate the pressure-temperature conditions of diagrams C-D and E, respectively.

(B) Oxygen fugacity-temperature relations for the Lesotho upper mantle, for the pressure-temperature relations given by the inflected geotherm from diagram A. The stippled field shows the oxygen fugacities and temperatures inferred from mineralogical studies (Eggler, 1983). Oxygen fugacities defined by synthetic mineral assemblages, considered relevant to upper mantle conditions, and in equilibrium with graphite/diamond are shown by curves. QFM: quartz - fayalite - magnetite; WM: wüstite - magnetite; EMO: orthopyroxene - magnesite - olivine; IW: iron - wüstite.

(C) Gas species compositions at 1023 °C and 50 kbar of a free C-O-H gas phase in equilibrium with graphite as a function of the oxygen fugacity (from Ryabchikov et al., 1981). Oxygen fugacity levels defined by the mineral assemblages in diagram B are indicated. (D) Molar H/(H + O) ratio of the gas compositions in diagram C as a function of oxygen fugacity.

(E) Gas species compositions at 1223 °C and 56 kbar of a free gas phase in equilibrium with graphite/diamond as a function of oxygen fugacity (from Eggler, 1983).

continental source regions for alkaline magmas and possibly subcontinental upper mantle in general.

The species composition of a gas phase existing at these temperatures and pressures as a function of the oxygen fugacity is shown in Fig. 6, C-E. At oxygen fugacities corresponding to those proposed from the natural upper mantle mineral assemblages the gas will essentially be a mixture of CO₂ and H₂O. Only at significantly lower oxidation levels (around iron - wüstite buffered oxygen fugacities) do more reduced gas species such as CH₄ take the place of CO₂.

The results in Fig. 6 indicate that if a gas exists in the upper mantle its

composition and oxidation state is most probably buffered to ($\text{H}_2\text{O} + \text{CO}_2$) -rich mixtures.

Although sulfur is not considered in the above presentation, results will not change significantly with addition of sulfur to the system. Calculations by Mitchell (1975), based on estimates of the sulfur fugacity from sulfide minerals found in kimberlites, show that for all reasonable sulfur fugacity levels any sulfur-bearing gas species in the C-O-H-S system such as H_2S , S_2 or SO_2 would only be present in concentrations 2-3 orders of magnitude less than H_2O , CO_2 , and CH_4 .

Gases in natural rocks of upper mantle derivation

Most of the evidence for the presence of deep gases and their composition is related to magmatic activity, most dramatically evidenced by the often explosive release of gases accompanying volcanic eruptions. The volcanic eruption itself is, however, only the final result of a series of events starting with partial melting of upper mantle, and the initial composition of any involved deep-seated gas phase may have been significantly changed in response to changes in temperature, pressure, composition and oxidation state of the magma during its ascent to the surface, and from mixing with ground and surface waters. Samples representing various intermediate stages in this sequence of events (Fig. 7) have, however, been preserved for study and a review of the data obtained from studies on such samples is presented below.

Compositions of volcanic and geothermal gases

Volcanic gases are generally considered to provide the most complete information on bulk gas composition because of the complete exsolution of all gases from silicate melts at atmospheric pressure.

Opportunities for field collection of volcanic gases are necessarily sporadic and involve serious access, collection and gas contamination problems. Available data are furthermore largely restricted to relatively quiet basaltic eruption thereby possibly biasing the obtained data. Gerlach and Nordlie (1975a; b; c) reviewed the then published volcanic gas compositions and attempted to remove the changes caused by atmospheric and meteoric water contamination, and reactions of the high-temperature gases with the sampling equipment from the analyses. Representative examples of such restored volcanic gas analyses are given in Table I.

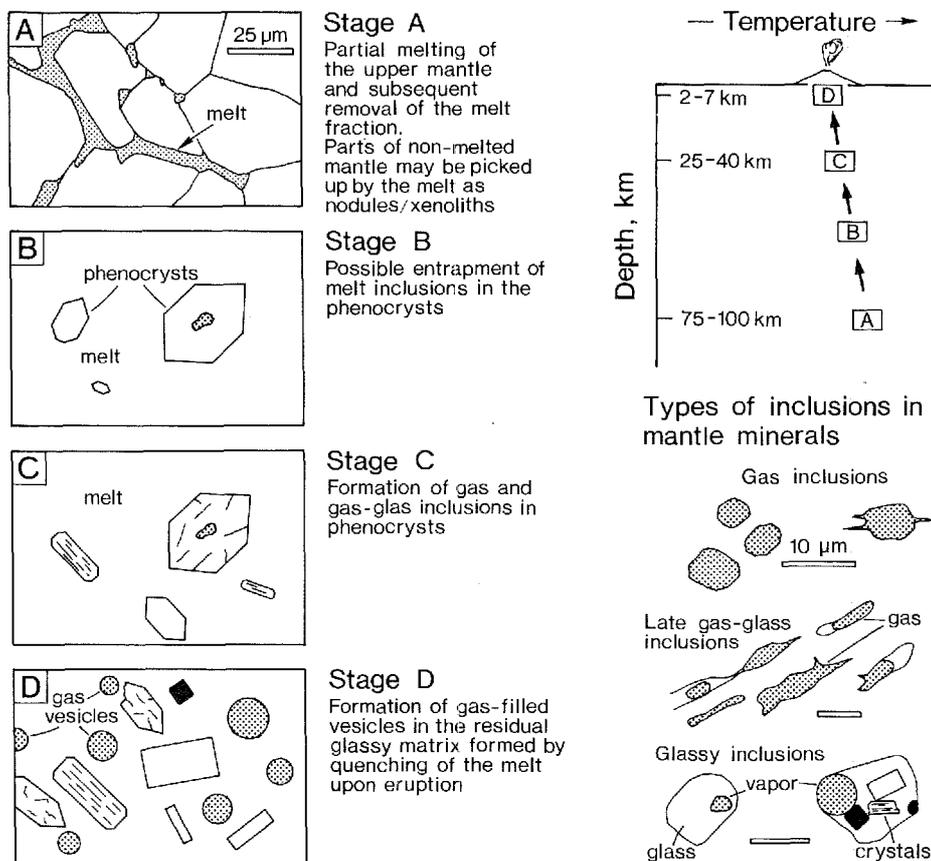


Fig. 7. Schematic representation of stages (A to D) of possible entrapment of gases in gas and melt inclusions in minerals in mantle-derived magmas, in gas vesicles and in the glassy matrix of volcanic rocks. Corresponding approximate depth and temperature conditions (to stages A-D) are sketched in the top-right figure. Sketches of the types and morphologies of inclusions in minerals from mantle nodules are shown in the lower-right part of the figure.

H_2O , CO_2 and SO_2 are the dominant species in all volcanic gases, generally in that order, and with other gases generally constituting only a few % of the volcanic gas. Thermodynamic calculations show all such gas compositions to be compatible with equilibration of the volcanic gas at oxygen fugacities generally constrained to values between those of the quartz-fayalite-magnetite and hematite-magnetite oxygen buffers, indicating effective buffering of the volcanic gas by the magma (Gerlach and Nordlie, 1975c).

TABLE I. Representative analyses of volcanic gases (mole %) released during eruptions of various basalt types

	(1)	(2)	(3)	(4)	(5)	(6)
H ₂ O	53.4	69.9	77.1	49.9	45.9	81.1
H ₂	0.7	1.6	1.6	0.5	1.6	2.8
CO ₂	33.3	17.8	11.7	21.7	45.6	9.3
CO	1.1	0.8	0.5	2.7	2.7	0.7
SO ₂	11.3	8.8	7.4	26.9	2.3	4.1
H ₂ S	0.05	1.0	0.9	0.2	1.4	0.9
S ₂	0.02	0.5	0.3	0.2	0.6	0.3
COS	—	0.02	—	—	0.1	—
HCl	0.1	—	0.4	—	—	0.8
T, °C	1175	1020	1020	1075	1020	1125

1. Kilauea, Hawaii, Tholeiitic basalts (from Gerlach, 1980a)
2. Erte'Ale, Ethiopia. 1971 eruption, basalts transitional between tholeiitic and alkaline (from Gerlach, 1979b)
3. Erte'Ale, Ethiopia. 1974 eruption, basalts transitional between tholeiitic and alkaline (from Gerlach, 1979b)
4. Mount Etna, Italy. 1970 eruption, alkali basalts (from Gerlach, 1979a)
5. Nyiragongo, Ethiopia. 1959 eruption, nephelinitic basalts (from Gerlach, 1980b)
6. Surtsey, Iceland. 1964-1967 eruptions, alkali olivine basalts (from Gerlach, 1980c)

Inspection of Table I and Fig. 8, where volcanic gases from various volcanoes are presented in terms of their C/S and C/H ratios, shows significant variations, both between different volcanoes and with time for an individual volcano (e.g. Erte'Ale, Table I). The overall distribution of volcanic gas compositions seen in Fig. 8, except for gases from the Nyiragongo lava lake, was proposed by Gerlach and Nordlie (1975a) possibly to reflect various degrees of degassing of basaltic magmas with a common initial gas composition, with C/H and C/S ratios of about 0.7-1.0 and 0.5-0.9, respectively. Alternatively, the distribution of gas compositions in Fig. 8 could represent true differences in gas compositions between different volcanoes and magma types (Sigvaldason, 1981).

Changes in volcanic gas composition with time for both Surtsey and Erte'Ale have been shown to be due to preferential degassing, especially of CO₂ (Gerlach, 1979b; 1980c). Such preferential degassing renders estimates of the initial bulk gas composition in magmas very difficult.

Degassing of species other than CO₂ from magma bodies has primarily been documented by studies of changes in compositions of gases in

fumaroles in areas with contemporaneous and repeated volcanic activity, such as the Krafla volcanic area in North Iceland. Although the majority of gases encountered in fumaroles on the basis of their chemical and isotopic compositions can be shown to have been leached from nearby non-magmatic rocks, both CO_2 and H_2 degassing from shallow magma chambers could be established (Oskarson, 1978; 1984). Outgassing of other minor species such as helium, radon and nitrogen are occasionally observed as well and may accompany degassing of the major species (CO_2) or occur independently (Oskarson, 1984).

Reduced gases in mid-ocean ridge hydrothermal systems

The discovery of active hydrothermal circulation systems at mid-oceanic ridges with exceptionally high contents of reduced gases such as CH_4 and H_2 , and accompanied by high helium contents with $^3\text{He}/^4\text{He}$ ratios about 8 times higher than atmospheric (e.g. Welhan and Craig, 1979) spurred the search for mantle derived gases in mid-ocean ridge hydrothermal systems. Studies of the isotopic composition of CH_4 and accompanying CO_2 and the overall chemical characteristics of these systems suggest that a common, abiogenic and deep origin for both the helium and the carbon in the reduced gases is highly likely (Welhan and Craig, 1982). Although the actual observed concentrations of the various gas species may reflect lower temperature equilibration of originally CO_2 and H_2O rich compositions via a Fisher-Tropsh type of reaction, it is also possible that the CH_4 is leached directly from underlying basalt together with H_2 , He and CO_2 . The carbon-isotopic composition observed in these mid-ocean ridge hydrothermal systems are very similar to those of

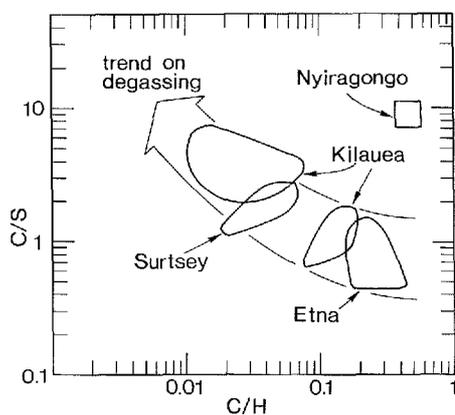
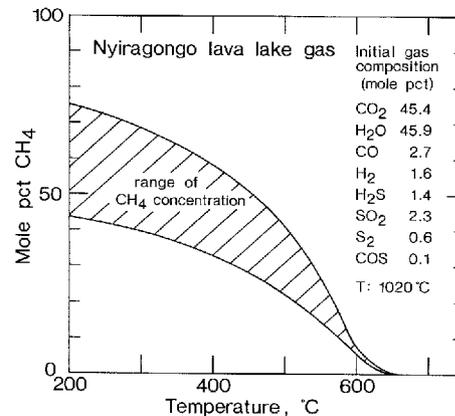


Fig. 8. Atomic C/S versus C/H relations of volcanic gases from various volcanoes as indicated in the figure (data from Gerlach, 1981; and references to Table I). Proposed evolution in gas composition with degassing of magma of a gas with an initial C/H and C/S ratio of about 0.7 and 0.5-1.0, respectively, is shown by the arrow (from Gerlach and Nordlie, 1975a).

Fig. 9. Concentration of CH_4 in Nyiragongo lava lake gas as a function of temperature, assuming equilibration of the gas at a pressure of 1 kbar and oxygen fugacities defined by the mineral assemblage quartz – fayalite – magnetite. The composition of the high-temperature volcanic gas is included in the figure. From Gerlach (1980b).



CO_2 found in vesicles in basalts and in the basaltic glass itself from mid-oceanic settings (Pineau et al., 1976; Moore et al., 1977; Pineau and Javoy, 1983; Des Marais and Moore, 1984).

High concentrations of reduced gases (up to 22% CH_4 and 77% CO_2) have been found in lake gases from Lake Kivu, Zaire, and stable isotope analyses (Deuser et al., 1973) have similarly indicated a deep, abiogenic origin for most of the carbon in this gas. The location of Lake Kivu in a continental rift environment (see Fig. 1) characterized by deep-going fault and fracture systems led Gold (1979) to consider the high CH_4 contents as possibly representing degassing from great depths. However, calculations by Gerlach (1980b) on volcanic gas compositions from the nearby Nyiragongo lava lake demonstrated that under conditions of continued buffering of such gas compositions by basaltic rock, at pressures corresponding to depths of about 5–6 km, gas compositions very rich in CH_4 would arise at temperatures below about 400 °C (Fig. 9). Therefore the high contents of CH_4 in Lake Kivu more probably represent lower-temperature equilibration of CO_2 -rich gases outgassed from underlying, partially molten magma chambers and not the degassing of CH_4 from depths.

Gases in vesicles, glassy matrix and melt inclusions in minerals from ocean-floor volcanics

Basaltic magmas erupted onto the sea-floor are generally considered to represent the best material for the study of the composition of the volcanic gas phase prior to eruption and gas saturation. Such melts are rapidly quenched under pressure thereby enabling the retention of dis-

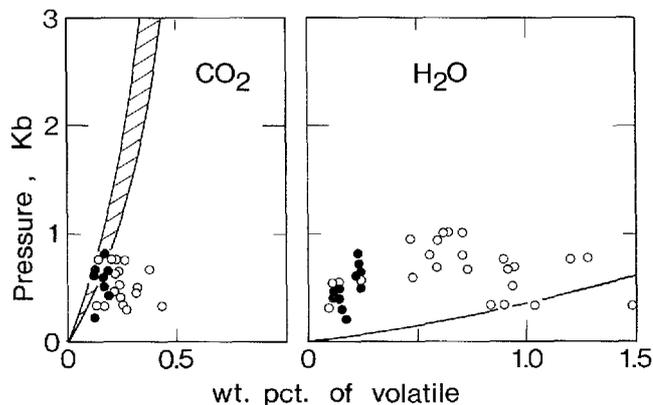
solved gases. Their chemical composition furthermore indicate that mid-ocean ridge basalts represent some of the least evolved mantle derived magmas erupted, and are thus closest in composition to the initially formed melt in the upper mantle.

The presence of vesicles in mid-ocean ridge basalt pillow lavas, however, indicates that these melts were already saturated with a gas phase at the time and pressure of eruption and quenching (Stage D, Fig. 7). Analyses of the gases in these vesicles show that CO_2 is the dominant gas, constituting more than 95 volume % of the gas in the vesicles, indicating that gas saturation of the melt prior to eruption was essentially due to CO_2 (Moore et al., 1977). In addition to CO_2 , a sulfur gas was originally present but later reacted to form sulfides now lining the vesicle walls. H_2O , SO_2 , O_2 and N_2 constitute the rest of the gas in these vesicles.

Analyses of the gases trapped in matrix glass from rims of the basaltic pillow lavas and of low vesicularity may reflect magma gas contents more closely. Data on the CO_2 and H_2O contents of such glassy matrix material are shown in Figs 10 and 11.

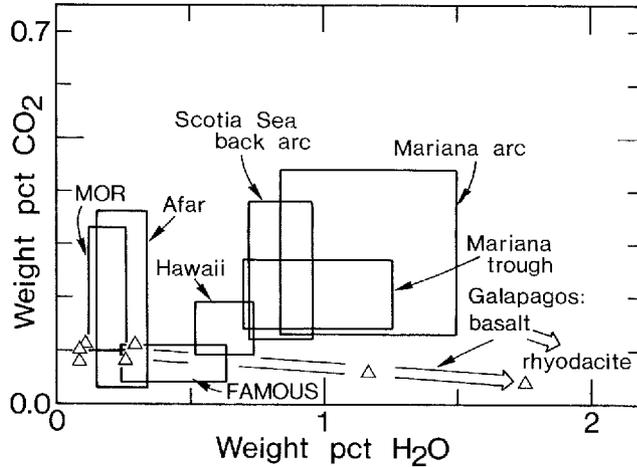
In Fig. 10 the CO_2 and H_2O contents are shown as a function of their quenching pressure (corresponding to their depth of eruption below the sea-floor) and further compared with the basalt- CO_2 and basalt- H_2O gas saturation curves, respectively. The data show that CO_2 saturation of the melts was probably attained in almost all samples prior to quenching.

Fig. 10. Concentration (in weight %) of CO_2 (left diagram) and H_2O (right diagram) in the glassy matrix of basalts quenched beneath the ocean floor shown as a function of the quenching pressure. Filled circles show data for mid oceanic ridge (MOR) basalts, open circles data for basalts from



other oceanic settings as discussed in the text. The results are shown in relation to the saturation curve for CO_2 and H_2O , respectively, in basalt. Sources of data: Killingley and Muenow (1975), Delaney et al. (1978), Garcia et al. (1979), Muenow et al. (1979), Chaigneau et al. (1980) and Muenow et al. (1980).

Fig. 11. Concentration (in weight %) of CO_2 versus H_2O in the glassy matrix of basalts from various geological locations and settings as indicated in the figure and discussed in the text. Data from Byers et al. (1984), and references to Fig. 10.



This suggests that the obtained CO_2 contents are the sum of CO_2 in the glass and vesicles, thereby preventing any estimate of the initial CO_2 content of the magmas. These basalts may have degassed a substantial amount of CO_2 and adjusted their actual CO_2 contents to the prevailing pressure prior to eruption and quenching. H_2O contents are, however, always lower than saturation values and may therefore be considered to express the actual H_2O contents of these melts.

In Fig. 11 a distinction between the gas contents in glassy matrix material from basalts of different tectonic setting and compositions has been made. Basalts from mid-ocean ridge areas (MOR, FAMOUS, Galapagos) and from Hawaii thus represent samples of magma believed to have suffered only a small degree of chemical evolution from their initial formation, whereas the remaining samples in Fig. 11 represent basalt types formed and erupted as a result of subduction of oceanic plate. The content of H_2O in the glassy matrix largely follows this distinction: H_2O contents of the glassy matrix in basalts from mid-ocean type of settings are significantly lower than that of subduction-related basaltic glasses. These latter glasses are also richer in elements such as Cl and F (e.g. Garcia et al., 1979), thus favoring models for formation of their magma involving addition of materials (e.g. volatiles) to their source regions from subducted oceanic crust. Such data furthermore show that gases in these tectonic setting may become recycled back to the upper mantle and that such recycling results in heterogeneities in the gas contents of the upper mantle.

The H₂O and CO₂ contents of a suite of glasses, ranging in composition from primitive mid-ocean ridge type basalts through andesite to rhyodacite, from the Galapagos rift have been emphasized in Fig. 11 in order to illustrate the relation between especially the H₂O content of the glasses and their chemical evolutionary stage. However, of most pertinence to deep gas compositions are the gas contents of the most primitive Galapagos basalt glasses (Byers et al., 1983). Total volatile contents for these glasses vary between 0.32 and 0.38 weight %, with 0.09-0.11 weight % H₂O and 0.08-0.11 weight % CO₂. As these glasses contain less than about 1 volume % gas vesicles their gas contents were considered to approximate actual gas contents for these melts. In addition to H₂O and CO₂, hydrocarbons as well as CO were detected, with the most abundant hydrocarbon being CH₄. CO contents were about 0.04 weight %, CH₄ contents < 0.01 weight %. The bulk composition of gases (H₂O: 36 wt. %, CO₂: 40 wt. %, CO: 20 wt. %, and CH₄: < 4 wt. %) in these glasses was considered by Byers et al. (1984) to represent approximately the gas composition of the source region for these melts. Assuming that these melts formed from 25% batch melting and highly incompatible behavior of the gas species, mantle source values of about 0.08 weight % total gases, and about 0.02 weight % of both CO₂ and H₂O are indicated. Such values indicate that very dry mantle may be involved in the generation of magmas along mid-ocean ridges.

Samples of the gas content of melt compositions at earlier stages (stages C and B in Fig. 7) during ascent of the magma are occasionally preserved as melt inclusions in phenocrysts. Analyses of such melt inclusions show that, whereas the mole percentage of CO₂ in the (residual) glassy matrix varies from about 4 to 38 mole % CO₂ in primitive basalt glasses, the mole percentage of CO₂ in melt inclusions in phenocrysts varies from about 45 to almost 100%. This trend towards higher H₂O contents with evolution of the magma is consistent with the evolution in gas composition of the melt to be expected from solubility data, if gas saturation occurred at some stage after the crystallization of the phenocrysts but prior to eruption. Samples indicative of such gas saturation of the melt prior to eruption are occasionally present in phenocrysts in the form of fluid inclusions.

Fluid inclusions in phenocrysts from lavas and in minerals from upper mantle xenoliths

CO₂ fluid inclusions in phenocrysts from lavas and in minerals from upper mantle xenoliths have been reported from more than 100 different localities worldwide (e.g. Roedder, 1984) and such inclusions appear to be typical of upper mantle derived materials, irrespective of their global tectonic environment. Representative examples of such CO₂-bearing fluid inclusions are sketched in Fig. 7.

In phenocrysts in lavas the CO₂ fluid is often seen to have been enclosed by the growing mineral together with silicate melt and this resulted in a mixed gas-glass inclusion at room temperature. Such simultaneous entrapment of fluid and melt shows that entrapment occurred at temperatures close to those of the gas-saturated solidus, that is temperatures around 1200 °C. The density of a CO₂ fluid in an inclusion will be identical to that of the free CO₂ phase coexisting with the mineral/magma at the temperature and pressure of entrapment. It is possible under the microscope to determine this density from the behavior of the enclosed CO₂ fluid. Knowledge of the density of a CO₂ fluid in an inclusion further enables the construction of the pressure-temperature path along which entrapment of that particular fluid inclusion occurred. If the pressure and temperature conditions indicated by the density (and chemical composition) of the entrapped fluid furthermore are compatible with the conditions of formation of the host mineral/rock, as inferred from other types of geological information, the entrapped fluid will provide information on the fluid conditions during mineral/rock formation.

Densities for CO₂ fluid inclusions in phenocrysts in lavas and in minerals from upper mantle xenoliths have been summarized in Fig. 12,B. Densities for the CO₂ fluid in gas-glass inclusions in phenocrysts have not been separated in Fig. 12,B but are generally less than 0.6-0.7 g/cm³. Such densities indicate entrapment of the CO₂ at pressures from a maximum of 3-5 kbar and down to a few hundred bars (Fig. 12,A). Although such pressures are far from upper mantle pressure values, the results indicate that CO₂ gas saturation in some magmas occurred at considerably greater depths than suggested from Fig. 8.

More common, however, and generally the type of CO₂ inclusions observed in minerals from mantle xenoliths are CO₂ fluid inclusions without any sign of the simultaneous entrapment of a silicate melt phase. Although the densities for such CO₂ inclusions extend to considerably higher values than those of the gas-glass inclusions (Fig. 12,B), estimates of the pressure of entrapment from these densities (to a maximum of

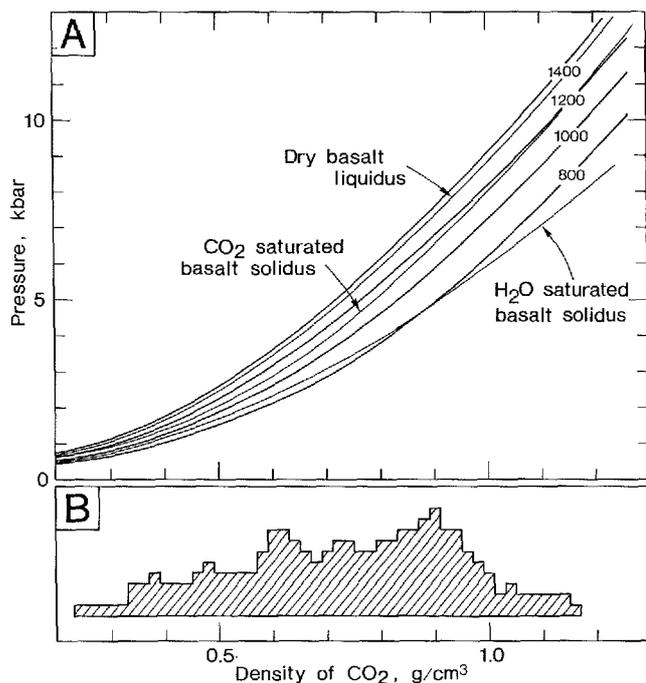


Fig. 12. Conditions for possible entrapment of CO₂ as gas inclusions in phenocrysts in lavas and in minerals from mantle derived nodules. (A) Possible entrapment pressures as function of the density of CO₂ in an inclusion at temperatures given by the selected four geotherms. Temperature-density-pressure curves corresponding to the basalt-CO₂ and basalt-H₂O saturated solidii and the liquidus curve for a gas-free basalt are shown for reference. Diagram based on data from Shmonov and Shmulovich (1974), Wyllie (1979), and Spera and Bergman (1980).

(B) Frequency diagram of CO₂ densities recorded for CO₂ inclusions in minerals

from mantle nodules and in phenocrysts from lavas from a large number of localities worldwide. Diagram summarized from data in Roedder (1965), Jones et al. (1983), Murck et al. (1978), Bergman and Dubessy (1984), Roedder (1984), and own data.

about 11-13 kbar; Fig. 12) are almost invariably significantly lower than the pressures indicated by the mineral assemblages in the xenoliths. The fact that no density higher than about 1.2 g/cm³ has been recorded for any CO₂ fluid inclusion in xenoliths, irrespective of depth of formation of the xenolith, suggests that, either the maximum densities recorded reflect the tensile strength of the host minerals and that any inclusion trapped at higher pressure will rupture during ascent of the xenolith to the surface and will be lost, or that the CO₂ fluid inclusions in minerals from mantle xenoliths do not reflect the presence of a free gas phase but owe their formation to some other mechanism.

Electron microscopic studies of μm -sized cavities, considered originally to have contained a gas, in minerals from mantle xenoliths suggest that cavity formation occurred in response to deformations in the mantle prior to incorporation of the xenolith material in the magma, and that cavity formation was due to nucleation and growth of bubbles on crystal

TABLE II. Representative analyses of the volatiles released from minerals on crushing in vacuum (mole %)

	(1)	(2)	(3)	(4)	(5)
H ₂ O	—	39.6	20.5	—	—
H ₂	7.0	19.4	5.0	0.2	54.0
CO ₂	16.1	15.5	0.3	—	18.5
CO	74.9	3.8	tr	50.4	—
CH ₄	1.2	3.7	64.2	23.0	2.2
C ₂ H ₆	tr	0.03	6.8	—	—
C ₃ H ₈	0.1	—	0.9	—	—
C _n H _{2n+2}	—	—	0.4	—	—
C _n H _{2n}	0.6	—	—	—	—
C ₃ HOH	—	1.0	—	—	—
CH ₃ CH ₂ OH	—	1.0	—	—	—
N ₂	0.1	15.9	2.8	26.4	25.2
Ar	—	0.1	0.1	—	—
He	—	1.0	0.03	—	0.005

1. Olivine from lherzolite nodule, Massif Central, France. (from Chaigneau, 1975)
2. Average of 7 natural diamonds, Arkansas, U.S.A. (from Melton and Giardini, 1975)
3. Nepheline from nepheline syenite, the illimaussaq intrusion, south Greenland (from Konnerup-Madsen et al., 1979)
4. Murchison C2 meteorite (from Andrawes and Gibson, 1979)
5. Average of 5 kimberlites, Obnazhennaya, U.S.S.R. (from Lutts et al., 1976)

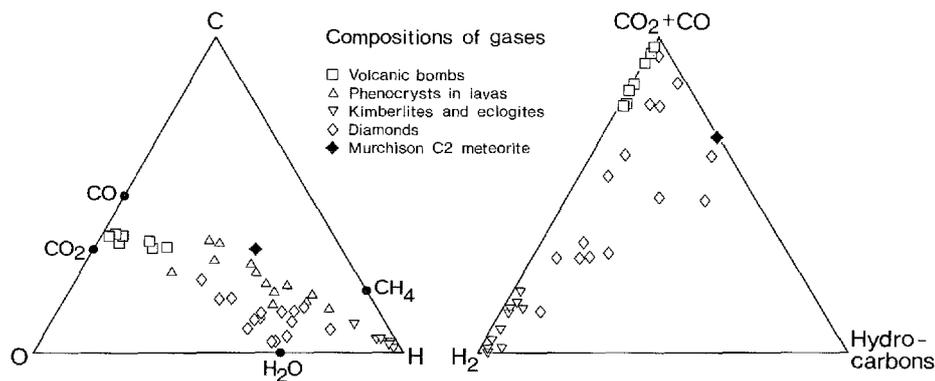
defects and these later collected to form larger cavities (Green and Radcliffe, 1975; Kirby and Green, 1980). It is, however, uncertain whether such a mechanism can be responsible also for the much larger (5–30 μm) CO₂ inclusions often observed in xenolith minerals. Alternative mechanisms for CO₂ fluid inclusion formation in such material include decarbonation of CO₂-bearing minerals occluded during growth of the host mineral, and precipitation of CO₂ directly from solid solution in the host mineral structure. The apparent lack of observations of decarbonation products in association with the CO₂ inclusions appears to favour the latter mechanism. Freund et al. (1980) reported appreciable amounts of carbon in xenolithic olivine, as atomic carbon in solid solution. Such carbon might exsolve at lower pressures (taking oxygen from ferric iron or other variable-valency elements) and form the CO₂ inclusions.

The composition of these CO₂ fluid inclusions is almost invariably pure CO₂, and all studies indicate that other gases are present in concentrations less than about 1 mole %. Only in one xenolith of wehrlite, with

an andesine-amphibole cross-cutting vein, have CO_2 inclusions containing 8–12 mole % CO been reported from the vein, whereas inclusions in the wehrlite nodule itself were pure CO_2 (Bergman and Dubessy, 1983).

Apart from the CO_2 gas compositions reported for high-density fluid inclusions, more complex gas compositions are frequently reported on other types of upper mantle derived materials such as kimberlites and diamonds. Representative examples of such analyses are included in Table II. Although the carbon isotopic signature of these gases clearly indicate their upper mantle origin, the general lack of recognizable gas inclusions in minerals from these materials when examined under the microscope, makes the actual location of evolved gases somewhat enigmatic. Gaseous constituents may be present in minerals in various forms (e.g. Bibby, 1982): (1) substitutionally or interstitially in the structure, (2) in mineral and melt inclusions or (3) as gaseous inclusions. For example, H, N and O are normally present in diamonds in the first form and in concentrations generally in the range 50–250 ppm although higher concentrations have been reported (Roedder, 1984). Therefore gas analyses such as those given in Table II may well represent the total amount of gases present at various locations in the minerals, and the actual bulk gas compositions reported may be the result of reactions among these gases during the extraction procedure preceding analyses. Thus, these analyses can not be taken as an expression of the gas species composition

Fig. 13. Compositions of gases in minerals and rocks of upper mantle derivation, as indicated in the figure, in terms of (left diagram) atomic C-O-H relations, and (right diagram) molecular ($\text{CO}_2 + \text{CO}$)- H_2 -total hydrocarbon contents and on a H_2O -free basis. References to data as to Fig. 12 and Table II, and Chemla et al. (1968), Krafft and Chaigneau (1980), and Vakin and Kutuyev (1980).



of any deep, upper mantle gas, although their compositions in terms of atomic C-O-H relations (Fig. 13) may be of significance. Except for the generally high contents of reduced gases in materials from the upper mantle such gases are distinguished from those previously discussed in containing high contents of N₂ (Table II). In both respects they resemble analyses obtained on gases from meteorites (Table II; and Yasinskaya, 1967; Fiéni et al., 1978).

Possible degassing along deep faults

High contents of hydrogen (> 3 volume % H₂) have been observed in soils and groundwaters from some areas with active faulting together with significantly increased helium fluxes and higher than normal ³He/⁴He ratios (Nagao et al., 1980). Such observations have been taken to document the possible degassing from deep sources of reduced gases together with helium, and in some fault areas other noble gases. Analyses of the D/H ratio indicate that the enrichment in hydrogen, however, is related to formation of molecular hydrogen from chemical reactions between groundwater and the surfaces of silicate minerals created by crushing during the fault movements (Kita et al., 1980). Experimental confirmation of the feasibility of such a mechanism has been given by Kita and Matsuo (1982). However, although no significant degassing of any major species is indicated to take place along fault zones, monitoring of gas compositions of soils and ground waters may prove useful in earthquake predictions.

Concluding comments on deep Earth gases

The actual concentration and composition of deep Earth gases is important because any such gases will represent a reservoir which may have yielded (parts of) the modern hydrosphere and atmosphere by degassing, and may have added nonbiological methane to the biologically produced hydrocarbon and oil deposits. Studies of deep-Earth gas compositions are, however, difficult because true samples from the upper mantle are difficult to separate unambiguously from samples contaminated by sedimentary and atmospheric gases.

The presently available evidence for degassing is mostly linked with helium. Enrichment in ³He has been established for a number of mantle

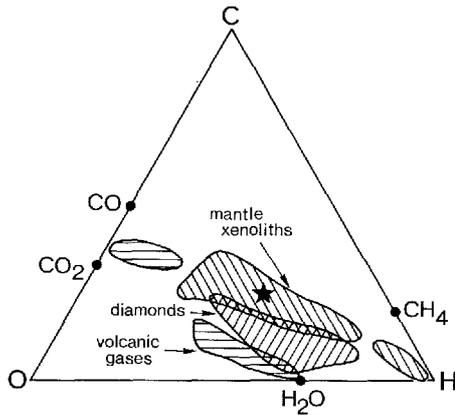


Fig. 14. Summary of the atomic C-O-H relations for gases in minerals and melts considered relevant to deep Earth gas compositions, as discussed in the text and summarized from data in previous figures. The star shows the bulk C-O-H composition estimated from gases in Galapagos basaltic glass matrix (Byers et al., 1984).

derived samples and, when coupled with observations on volcanic and fumarolic gases, CO₂ and possibly some more reduced gases in mid-ocean ridge hydrothermal systems, gas vesicles and the glassy matrix of ocean-floor basalts, clearly demonstrate that volatiles from deep sources are continuously being added to the Earth's surface. Attempts to quantify this addition of gases to the Earth's atmosphere have been made on the basis of gases in diamonds (Melton and Giardini, 1982) and from noble and rare gas isotope systematics in ocean-floor basalts (Kyser and Rison, 1982; Ozima et al., 1983) but the results vary considerably and more reliable estimates must await further studies.

Estimates of the actual major gas species composition of any deep gas may be attempted on the basis of data from compositions of volcanic gases, gases in rapidly quenched ocean-floor volcanics, or from gas compositions of gases trapped by some mechanism in minerals in mantle xenoliths (Fig. 14). Such estimates vary considerably, both in terms of atomic proportions of C, O, and H, and in terms of the corresponding major gas species compositions. Part of the discrepancy between the various sets of data probably reflects differences in oxidation states during formation of the different materials, but other factors such as solid state diffusion of gaseous constituents, either in elemental form or as gas species, may be important.

Although the possible existence of a free gas phase in some parts of the upper mantle is predicted from experimental and theoretical studies, and will essentially be a mixture of CO₂ and H₂O, actual gas concentrations suggested for some source regions from the gas contents of primitive ocean-floor basalts indicate the possible storage of both C, H and O in

minerals as impurities in the crystal structure or forming solid solutions. Especially the state of carbon in the Earth's mantle is crucial to any hypothesis regarding deep Earth gas compositions. If carbon is present in largely elemental form, much carbon could migrate as a solid phase through the mineral lattices and produce CH₄ and CO₂ only when reducing or oxidizing conditions were encountered (Freund, 1981; Duba and Shankland, 1982). If such mechanisms are operative on an important scale, earlier perceived ideas regarding the actual form of gases in the Earth's mantle and of the processes operating may be highly misleading.

ACKNOWLEDGEMENTS. *I am grateful for the constructive criticism and comments by J. Bailey and H. Sørensen. The financial support provided by the Danish Natural Science Research Council (J.no. 11-3837) during the opening stages of this study is gratefully acknowledged.*

References

- Aines, R. D. and Rossman, G. R. 1984. Water in minerals? A peak in the infrared. *Jour. Geophys. Res.*, 89: 4059-4071.
- Anderson, O. L. 1981. A decade of progress in earth's internal properties and processes. *Science*, 213: 76-82.
- Andrawes, F. F. and Gibson, E. K. 1979. Release and analysis of gases from geological samples. *Am. Mineral.*, 64: 453-463.
- Arculus, R. J. and Delano, J. W. 1980. Implications for the primitive atmosphere on the oxidation state of the earth's upper mantle. *Nature*, 288: 72-74.
- Bergman, S. C. and Dubessy, J. 1984. CO₂-CO fluid in a composite peridotite nodule: implications for oxygen barometry. *Contrib. Mineral. Petrol.*, 85: 1-13.
- Bibby, D. M. 1982. Impurities in natural diamond. In P. A. Throver (Ed.), *Chemistry and Physics of Carbon*, Vol. 18, 1-91. Marcel Dekker, New York.
- Boyd, F. R. 1973. A pyroxene geotherm. *Geochim. Cosmochim. Acta*, 37: 2533-2546.
- Bundy, F. P., Bovenderk, H. P., Strong, H. M. and Wentorf, R. H. 1961. Diamond-graphite equilibrium line from growth and graphitization of diamond. *Jour. Chem. Phys.*, 35: 383-396.
- Burnham, C. W. and Davis, N. F. 1974. The role of H₂O in silicate melts. II. Thermodynamic and phase relations in the system NaAlSi₃O₈-H₂O to 10 kilobars, 700° and 1100 °C. *Am. Jour. Sci.*, 274: 902-940.
- Byers, C. D., Muenow, D. W. and Garcia, M. O. 1983. Volatiles in basalts and andesites from the Galapagos spreading center, 85° and 86°N. *Geochim. Cosmochim. Acta*, 47: 1551-1558.
- Byers, C. D., Christie, D. M., Muenow, D. W. and Sinton, J. M. 1984. Volatile contents and ferric-ferrous ratios of basalts, ferrobalt, andesite and rhyodacite glasses from the Galapagos 95.5°W propagating rift. *Geochim. Cosmochim. Acta*, 48: 2239-2246.

- Chaigneau, M. 1975. Essai sur la composition des gaz magmatiques profonds. *Bull. Volcanol.*, 39: 1-13.
- Chaigneau, M., Hékinian, R. and Cheminée, J. L. 1980. Magmatic gases extracted and analysed from ocean floor volcanics. *Bull. Volcanol.*, 43: 241-253.
- Chemla, M., Toray, J. C., Brousse, R. and Causse, C. 1968. Abondance relative des gaz inclus dans quelques minéraux magmatiques d'origine profonde. In L. H. Ahrens (Ed.), *Origin and Distribution of the Elements*, 971-979. Pergamon Press, London.
- Cogley, J. G. and Henderson-Sellers, A. 1984. The origin and earliest state of the Earth's hydrosphere. *Rev. Geophys. Space Phys.*, 22: 131-170.
- Dawson, J. B. 1980. *Kimberlites and Their Xenoliths*. 252 p. Springer Verlag, Berlin.
- Dawson, J. B. 1981. The nature of the upper mantle. *Min. Mag.*, 44: 1-18.
- Dawson, J. B. and Smith, J. V. 1982. Upper-mantle amphiboles: a review. *Min. Mag.*, 45: 35-46.
- Delaney, J. R., Muenow, D. W. and Graham, D. G. 1978. Abundance and distribution of water, carbon and sulfur in the glassy rims of submarine pillow basalts. *Geochim. Cosmochim. Acta*, 42: 581-594.
- Des Marais, D. J. and Moore, J. G. 1984. Carbon and its isotopes in mid-oceanic basaltic glasses. *Earth Planet. Sci. Letts*, 69: 43-57.
- Deuser, W. G., Degens, E. T. and Harvey, G. R. 1973. Methane in lake Kivu: new data bearing on its origin. *Science*, 181: 51-53.
- Duba, A. G. and Shankland, T. J. 1982. Free carbon & electrical conductivity in the Earth's mantle. *Geophys. Res. Letts*, 9: 1271-1274.
- Eggler, D. H. 1983. Upper mantle oxidation state: evidence from olivine-orthopyroxene-ilmenite assemblages. *Geophys. Res. Letts*, 10: 365-368.
- Eggler, D. H. and Kadik, A. A. 1979. The system $\text{NaAlSi}_3\text{O}_8\text{-H}_2\text{O-CO}_2$ to 20 kbar pressure: I. Compositional and thermodynamic relations of liquids and vapors coexisting with albite. *Am. Mineral.*, 64: 1036-1048.
- Eggler, D. H., Baker, D. R. and Wendlandt, R. F. 1980. f_{O_2} of the assemblage graphite-enstatite-forsterite-magnesite: experiment and application to mantle f_{O_2} and diamond formation. *Geol. Soc. Am. Abstr. Programs*, 12: 420.
- Eggler, D. H., Mysen, B. O., Hoering, T. C. and Holloway, J. R. 1979. The solubility of carbon monoxide in silicate melts at high pressures and its effect on silicate phase relations. *Earth Planet. Sci. Letts*, 43: 321-330.
- Fiéni, C., Bouret-Denise, M., Pellas, P. and Touret, J. 1978. Aqueous fluid inclusions in feldspars and phosphates from Peetz chondrite (Abstract). *Meteoritics*, 13: 460-461.
- Freund, F. 1981. Mechanism of the water and carbon dioxide solubility in oxides and silicates and the role of O^- . *Contrib. Mineral. Petrol.*, 76: 474-482.
- Freund, F., Kathrein, H., Wengeler, H., Knobel, R. and Heinen, 1980. Carbon in solid solution in forsterite – a key to the untractable nature of reduced carbon in terrestrial and cosmogenic rocks. *Geochim. Cosmochim. Acta*, 44: 1319-1333.
- Garcia, M. O., Liu, N. W. K. and Muenow, D. W. 1979. Volatiles in submarine volcanic rocks from the Mariana island arc and trough. *Geochim. Cosmochim. Acta*, 43: 305-312.
- Gerlach, T. M. 1979a. Evaluation and restoration of the 1970 volcanic gas analyses from Mount Etna, Sicily. *Jour. Volcanol. Geotherm. Res.*, 6: 165-178.
- Gerlach, T. M. 1979b. Investigation of volcanic gas analyses and magma outgassing from Erte'Ale lava lake, Ethiopia. *Jour. Volcanol. Geotherm. Res.*, 7: 415-441.

- Gerlach, T. M. 1980a. Evaluation of volcanic gas analyses from Kilauea volcano. *Jour. Volcanol. Geotherm. Res.*, 7: 295-317.
- Gerlach, T. M. 1980b. Chemical characteristics of the volcanic gases from Nyiragongo lava lake and the generation of CH₄-rich fluid inclusions in alkaline rocks. *Jour. Volcanol. Geotherm. Res.*, 8: 118-121.
- Gerlach, T. M. 1980c. Evaluation of volcanic gas analyses from Surtsey volcano, Iceland, 1964-1967. *Jour. Volcanol. Geotherm. Res.*, 8: 191-198.
- Gerlach, T. M. 1981. Restoration of new volcanic gas analyses from basalts of the Afar region: further evidence of CO₂-degassing trends. *Jour. Volcanol. Geotherm. Res.*, 10: 83-91.
- Gerlach, T. M. and Nordlie, B. E. 1975a. The C-O-H-S gaseous system, Part I: composition limits and trends in basaltic gases. *Am. Jour. Sci.*, 275: 353-376.
- Gerlach, T. M. and Nordlie, B. E. 1975b. The C-O-H-S gaseous system, part II: Temperature, atomic composition and molecular equilibria in volcanic gases. *Am. Jour. Sci.*, 275: 377-394.
- Gerlach, T. M. and Nordlie, B. E. 1975c. The C-O-H-S gaseous system, part III: magmatic gases compatible with oxides and sulfides in basaltic magmas. *Am. Jour. Sci.*, 275: 395-410.
- Giardini, A. A. and Melton, C. E. 1975. The nature of cloud-like inclusions in two Arkansas diamonds. *Am. Mineral.*, 60: 931-933.
- Gold, T. 1979. Terrestrial sources of carbon and earthquake outgassing. *Jour. Petroleum Geol.*, 1: 3-19.
- Green, H. W. and Radcliffe, S. V. 1975. Fluid precipitates in rocks from the Earth's mantle. *Geol. Soc. Am. Bull.*, 86: 846-852.
- Jones, A. P., Smith, J. V. and Dawson, J. B. 1983. Glasses in mantle xenoliths from Olmani, Tanzania. *Jour. Geol.*, 91: 167-178.
- Kadik, A. A. and Lukanin, O. A. 1973. The solubility-dependent behavior of water and carbon dioxide in magmatic processes. *Geochemistry Internat.*, 34: 115-129.
- Kesson, S. E. and Holloway, J. R. 1974. The generation of N₂-CO₂-H₂O fluids for use in hydrothermal experimentation II. Melting of albite in a multispecies fluid. *Am. Mineral.*, 59: 598-603.
- Killingley, J. S. and Muenow, D. W. 1975. Volatiles from Hawaiian submarine basalts determined by dynamic high temperature mass spectrometry. *Geochim. Cosmochim. Acta*, 49: 1467-1473.
- Kirby, S. H. and Green II, H. W. 1980. Dunite xenoliths from Hualalai volcano: evidence for mantle diapiric flow beneath the island of Hawaii. *Am. Jour. Sci.*, 280-A: 550-575.
- Kirsten, T. 1968. Incorporation of rare gases in solidifying enstatite melts. *Jour. Geophys. Res.*, 73: 2807-2810.
- Kita, I. and Matsuo, S. 1982. H₂ generation by reaction between H₂O and crushed rock: an experimental study on H₂ degassing from the active fault zone. *Jour. Geophys. Res.*, 87: 10789-10795.
- Kita, I., Matsuo, S., Wakita, H. and Nakamura, Y. 1980. D/H ratios of H₂ in soil gases as an indicator of fault movements. *Geochem. Jour.*, 14: 317-320.
- Konnerup-Madsen, J., Larsen, E. and Rose-Hansen, J. 1979. Hydrocarbon-rich fluid inclusions in minerals from the alkaline Ilímaussaq intrusion, south Greenland. *Bull. Minéral.*, 102: 642-653.

- Krafft, M. and Chaigneau, M. 1980. Les gaz occlus dans les bombes volcaniques de l'activité du Piton de la Fournaise en 1975-1976 (Ile de la Réunion). *Bull. Volcanol.*, 43: 225-232.
- Kyser, T.K. and Rison, W. 1982. Systematics of rare gas isotopes in basic lavas and ultramafic xenoliths. *Jour. Geophys. Res.*, 87: 5611-5630.
- Lutts, B. G., Petersil'ye, I. A. and Karzhavin, V. K. 1976. Composition of gases in rocks of the upper mantle. *Doklady Akad. Nauk. SSSR*, 226: 197-200.
- MacDonald, G. J. 1983. The many origins of natural gas. *Jour. Petroleum Geol.*, 5: 341-362.
- Mathez, E. A. 1984. Carbon abundances in mantle minerals determined by nuclear reaction analysis. *Geophys. Res. Letts*, 11: 947-950.
- Matson, D. W. and Muenow, D. W. 1984. Volatiles in amphiboles from xenoliths, Vulcans Throne, Grand Canyon, Arizona, USA. *Geochim. Cosmochim. Acta*, 45: 1629-1636.
- Melton, C. E. and Giardini, A. A. 1974. The composition and significance of gas released from natural diamonds from Africa and Brazil. *Am. Mineral.*, 59: 775-782.
- Melton, C. E. and Giardini, A. A. 1975. Experimental results and a theoretical interpretation of gaseous inclusions found in Arkansas natural diamonds. *Am. Mineral.*, 60: 413-417.
- Melton, C. E. and Giardini, A. A. 1982. The evolution of the Earth's atmosphere and oceans. *Geophys. Res. Letts*, 9: 579-582.
- Mitchell, R. H. 1975. Theoretical aspects of gaseous and isotopic equilibria in the system C-H-O-S with application to kimberlite. In L. H. Ahrens, J. B. Dawson, A. R. Duncan and A. J. Erlank (Eds), *Physics and Chemistry of the Earth*, Vol. 9, 903-915. Pergamon Press, New York.
- Moore, J. G., Batchelder, J. N. and Cunningham, C. G. 1977. CO₂-filled vesicles in mid-ocean basalt. *Jour. Volcanol. Geotherm. Res.*, 2: 309-327.
- Muenow, D. W., Graham, D. G., Liu, N. W.K. and Delaney, J. R. 1979. The abundance of volatiles in Hawaiian tholeiitic submarine basalts. *Earth Planet. Sci. Letts*, 42: 71-76.
- Muenow, D. W., Liu, N. W.K., Garcia, M. O. and Saunders, A. D. 1980. Volatiles in submarine volcanic rocks from the spreading axis of the East Scotia Sea backarc basin. *Earth. Planet. Sci. Letts*, 47: 272-278.
- Murck, B. W., Burruss, R. C. and Hollister, L. S. 1978. Phase equilibria in fluid inclusions in ultramafic xenoliths. *Am. Mineral.*, 63: 40-46.
- Mysen, B. O. 1976. The role of volatiles in silicate melts: solubility of carbon dioxide and water in feldspar, pyroxene, and feldspatoid melts to 30 kbar and 1625 °C. *Am. Jour. Sci.*, 276: 969-996.
- Mysen, B. O. and Boettcher, A. L. 1975. Melting of hydrous mantle: I. Phase relations of natural peridotite at high pressures and temperatures with controlled activities of water, carbon dioxide, and hydrogen. *Jour. Petrology*, 16: 520-548.
- Mysen, B. O., Egger, D. H., Seitz, M. G. and Holloway, J. R. 1976. Carbon dioxide in silicate melts and crystals. Part I. Solubility measurements. *Am. Jour. Sci.*, 276: 455-479.
- Nagao, K., Takaoka, N., Wakita, H., Matsuo, S. And Fujii, N. 1980. Isotopic compositions of rare gases in the Matsushiro earthquake fault region. *Geochem. Jour.*, 14: 63-69.

- Nordlie, B. E. 1971. The composition of the magmatic gas of Kilauea and its behavior in the near surface environment. *Am. Jour. Sci.*, 271: 417-432.
- Olafsson, M. 1980. Partial melting of peridotite in the presence of small amounts of volatiles, with special reference to the low-velocity zone. M.Sc. Thesis, 59 p. Pennsylvania State Univ., USA.
- Oskarson, N. 1978. Effect of magmatic activity on fumarole gas composition in the Námafjall-Krafla volcanic center, N-Iceland. *Nordic Volcanol. Inst., Prof. Pap.* 7803: 1-81.
- Oskarson, N. 1984. Monitoring of fumarole discharge during the 1975-1982 rifting in Krafla volcanic center, Iceland. *Jour. Volcanol. Geotherm. Res.*, 22: 97-122.
- Ozima, M., Zashu, S. and Nitoh, O. 1983. $^3\text{He}/^4\text{He}$ ratio, noble gas abundance and K-Ar dating of diamonds – An attempt to search for records of early terrestrial history. *Geochim. Cosmochim. Acta*, 47: 2217-2224.
- Pineau, F. and Javoy, M. 1983. Carbon isotopes and concentrations in mid-oceanic ridge basalts. *Earth Planet. Sci. Letts*, 62: 239-257.
- Pineau, F., Javoy, M. and Bottinga, Y. 1976. $^{13}\text{C}/^{12}\text{C}$ ratios of rocks and inclusions in popping rocks of the mid-atlantic ridge and their bearing on the problem of isotopic composition of deep-seated carbon. *Earth. Planet. Sci. Letts*, 29: 413-421.
- Poreda, R. J. and Basu, A. R. 1984. Rare gases, water, and carbon in kaersutites. *Earth Planet. Sci. Letts*, 69: 58-68.
- Rai, C. S., Sharma, S. K., Muenow, D. W., Matson, D. W. and Byers, C. D. 1983. Temperature dependence of CO_2 solubility in high pressure quenched glasses of diopside composition. *Geochim. Cosmochim. Acta*, 47: 953-958.
- Roedder, E. 1965. Liquid CO_2 in olivine-bearing nodules and phenocrysts in basalts. *Am. Mineral.*, 50: 1746-1786.
- Roedder, E. 1984. Fluid Inclusions. *Reviews in Mineralogy*, Vol. 12, 644 p. Mineralogical Soc. America.
- Ryabchikov, I. D., Green, D. H., Wall, W. J. and Brey, G. P. 1981. The oxidation state of carbon in the reduced-velocity zone. *Geochemistry Internat.*, 18: 148-158.
- Shmonov, V. M. and Shmulovich, K. I. 1974. Molal volumes and equation of state of CO_2 at temperatures from 100° to 1000 °C and pressures from 2000 to 10.000 bars. *Doklady Akad. Nauk. SSSR*, 217: 935-938.
- Sigvaldason, G. E. 1981. Fluids in volcanic and geothermal systems. In D. T. Rickard and F. E. Wickman (Eds), *Chemistry and Geochemistry of Solutions at high Temperatures and Pressures. Physics and Chemistry of the Earth*, Vols 13 & 14, 179-191. Pergamon Press, New York.
- Spera, F. J. and Bergman, S. C. 1980. Carbon dioxide in igneous petrogenesis. I. *Contrib. Mineral. Petrol.*, 74: 55-66.
- Vakin, E. A. and Kutyev, F. S. 1980. Depth of origin of the gas content of thermal waters. *Bull. Volcanol.*, 43: 85-93.
- Welhan, J. A. and Craig, H. 1979. Methane and hydrogen in East Pacific Rise hydrothermal fluids. *Geophys. Res. Letts*, 6: 829-831.
- Welhan, J. A. and Craig, H. 1982. Abiogenic methane in mid-ocean ridge hydrothermal fluids. In W. J. Gwilliam (Ed.), *Deep Source Gas Workshop Technical Proceedings*, 122-129. Morgantown, USA.
- Wendlandt, R. F. 1982. Sulfide saturation of basalt and andesite melts at high pressures and temperatures. *Am. Mineral.*, 67: 877-885.

- Wyllie, P. J. 1971. *The Dynamic Earth*. 416 p. Wiley & Sons, Inc., New York.
- Wyllie, P. J. 1979. Magmas and volatile components. *Am. Mineral.*, 64: 469-500.
- Wyllie, P. J. and Huang, W. L. 1975. Peridotite, kimberlite, and carbonatite explained in the system CaO-MgO-SiO₂-CO₂. *Geology*, 3: 621-624.
- Wyllie, P. J. and Huang, W. L. 1976. Carbonation and melting reactions in the system CaO-MgO-SiO₂-CO₂ at mantle pressures with geophysical and petrological applications. *Contrib. Mineral. Petrol.*, 54: 79-107.
- Yasinskaya, A. A. 1967. Inclusions in stony meteorites. *Mineralog. Sbornik L'vov Gos. Univ.*, 21: 278-281.

