A COMPARISON OF THE DETERMINATION OF CLOSURE PHASE IN OPTICAL INTERFEROMETRY WITH FULLY FILLED APERTURES AND NON-REDUNDANT APERTURE MASKS

A.C.S. Readhead

Owens Valley Radio Observatory
California Institute of Technology

1. INTRODUCTION

The recent successes in diffraction limited optical imaging from the ground by Weigelt et al. (1986) and Baldwin et al. (1986) rely on the use of the closure phase - i.e. the sum of the visibility phases around a closed loop of (usually three) baselines (Jennison 1953, 1958). The closure phase has been used for the last dozen years in VLBI for making diffraction limited images in situations where the observed phase of the the incoming wavefront is badly corrupted by propagation and instrumental effects (Rogers et al. 1974; Readhead and Wilkinson 1978; Readhead 1980; Cornwell and Wilkinson 1981; Schwab 1980; Pearson and Readhead 1984).

The approach adopted by Baldwin et al. is a direct carry-over of the methods of VLBI - a non-redundant mask is used to transmit light from a few patches on the primary mirror of a large telescope, each patch being coherently illuminated and therefore analogous to the independent telescopes used in VLBI. The approach of Weigelt et al. makes use of the whole aperture. They form the bispectrum, or triple product, from speckle images and use this to make the image (Lohmann et al. 1983). The bispectrum is the product of the complex visibility function at three points in the (u,v) plane which form a closed set of baseline vectors, and the phase of the triple product is therefore the closure phase (Cornwell 1987) plus a noise term which is introduced by the redundant baselines, as described below.

The object structure information obtained by these two approaches is therefore the visibility amplitude and the closure phase. What makes these methods different and more powerful than previous techniques employed in optical interferometry is the use of the closure phase.

Since the closure phase is the basic new piece of information that is used in <u>both</u> of these methods, we will avoid the unfortunate and misleading apellations "bi-spectrum" and "closure phase" which are sometimes used to distinguish between these two approaches. The real distinction is clearly between the use of 'fully filled apertures' and 'non-redundant masks', and we therefore use these terms.

This paper presents an analysis of the effects of the redundant baselines on closure phase measurements in the 'fully filled aperture' approach for the case of many photons , i.e. ignoring the effects of photon noise, which provides some useful insights into the effects of redundant baselines which exist in both the high- and low-photon rate cases, and which therefore must be taken into account in the critical case of low photon rates.

2. THE MEASUREMENT OF CLOSURE PHASE IN OPTICAL INTERFEROMETRY

In the following discussion we first consider (in §2.1) the measurement of closure phase with a non-redundant 3-hole aperture mask, i.e. the technique used by Baldwin et al. We determine the effect on the closure phase of integrating for longer than the coherence time. We show that the use of the closure phase, in addition to making it possible to determine the object visibility phase, actually extends the effective coherence time for individual observations. It is important to understand this distinction - the closure phase does, of course, effectively enable us to make coherent observations for periods of arbitrary duration (typically 12 hours in VLBI) by using the object as its own phase reference; in addition the coherence of individual raw observations is extended, as shown below, because of the cancellation of some phase terms.

The results of this analysis in the time domain are then applied (in §2.2) in the spatial domain to the case of fully filled apertures.

2.1 Non-Redundant Masks

We begin by considering the formation of an intensity fringe pattern by interference of light from three elementary areas i,j,k. Let the amplitudes and phases of the intensity patterns corresponding to the three intersecting fringe patterns thus formed be represented by the vectors $A_{ij}e^{i\phi_{ij}}$, $A_{jk}e^{i\phi_{jk}}$ and $A_{ki}e^{i\phi_{ki}}$ on baselines ij,jk and ki, respectively.

Consider an observation for which the integration time, τ, is significantly longer than the coherence time, τ_{coh}. For simplicity, and without loss of generality, we assume that the fringe pattern is frozen for intervals $\tau_o \sim \tau_{coh}$, with random phase changes of order one radian at each of the three apertures between each interval τ_o, and that $\tau = n\tau_o$. We will examine the intensity pattern obtained in such a situation. The Fourier Transform of the intensity pattern arising from baseline ij for an observation of duration τ is simply the vector sum:

$$A_{ij_1}e^{i(\phi_{ij_1})} + A_{ij_2}e^{i(\phi_{ij_2})} + \ldots A_{ij_n}e^{i(\phi_{ij_n})};$$

where the subscripts 1,2 . . . n refer to the 1st, 2nd . . . nth interval of duration τ_o. The triple product on the baselines ij, jk and ki is:

$$\left(A_{ij_1}e^{i\phi_{ij_1}} + \ldots + A_{ij_n}e^{i\phi_{ij_n}}\right)\left(A_{jk_1}e^{i\phi_{jk_1}} + \ldots + A_{jk_n}e^{i\phi_{jk_n}}\right)\times$$
$$\left(A_{ki_1}e^{i\phi_{ki_1}} + \ldots + A_{ki_n}e^{i\phi_{ki_n}}\right) \quad \ldots\ldots\ldots (1)$$

This contains terms of three types (Readhead et al. 1987):

1) There are n combinations of complex visibilities over the same interval, which give us the closure phase, and which we will call 'terms of the first kind'.

2) There are 3n(n-1) combinations involving cross terms from two intervals, which we call 'terms of the second kind'. These terms of the second kind can be combined in complementary pairs for which the resultants either have phase equal to the closure phase or they differ from the closure phase by π. Combining 3n(n-1)/2 of these pairs results in a vector of rms length $[3n(n-1)]^{1/2}$ at angle ϕ_{ijk} or $\phi_{ijk}+\pi$. When the terms of the second kind are added to those of the first kind they therefore produce no change in the closure phase or they produce a change of π.

3) There are n(n-1)(n-2) combinations of three intervals. These terms have random phases, and therefore they add incoherently. They introduce a noise term into the triple product, and corrupt the closure phases of the triple product. We call these 'terms of the third kind'.

For the sake of the present discussion we will assume that all the amplitudes, A_{ij}, are equal and set these equal to unity, i.e. we assume that the scintillation is negligible. The bispectrum then consists of a coherent (constant phase) term of length n, and phase ϕ_{ijk}; a term of phase ϕ_{ijk} or $\phi_{ijk}+\pi$ and rms length $[3n(n-1)]^{1/2}$; and a term of random phase and rms length $[n(n-1)(n-2)]^{1/2}$. Note that if only two intervals are considered, i.e. if n = 2, there are no terms of the third kind, so that the noise in the closure phase is not increased by the longer integration time, but it will be wrong by π 21% of the time.

Apart from this possible uncertainty of π in the closure phase, the signal-to-noise-ratio is $n/[n(n-1)(n-2)]^{1/2}$, so that for a signal-to-noise-ratio of unity, corresponding roughly to an rms error in the measured closure phase of 1 radian, $n_c = 3.41$. While for an rms closure phase error of 104°, the rms value of a random variable with uniform distribution between $-\pi$ and π, $n_c \sim 6$. Thus the use of the closure phase effectively extends the coherence time by a significant factor, the reason for this extension in coherence time being the cancellation of the terms of the second kind.

In order to measure the closure phase we have to take many exposures and add the bispectra. Since the closure phase is preserved, the terms of the first kind add coherently, while those of the second kind add incoherently and affect the amplitude of the coherent signal. Thus, after summing N exposures we have the situation illustrated in Figure 1. Note that the amplitude of the fixed phase term is $Nn \pm [N3n(n-1)]^{1/2}$; while the rms amplitude of the incoherent random phase term is $[Nn(n-1)(n-2)]^{1/2}$. This shows that it is possible to extract closure phase information from observations which extend to much longer than τ_{coh}, but that the number of frames needed to reduce the rms error in the closure phase to 1/x radians is $N = x^2(n-3+2/n)$. Ideally, of course, one would observe for $\tau < \tau_{coh}$ and so eliminate these noise terms altogether.

2.2 Fully Filled Apertures

The above analysis can also be applied in the spatial domain. Consider the intensity pattern produced by a fully filled aperture. There are many identical triangles of elementary coherence areas, and the triple product of the complex visibility of these redundant triangles is again given by (1). In the case of 'frozen turbulence', i.e. under the Taylor hypothesis, there is no difference between combining the interference fringes from a single triplet of elementary areas over a period of time and combining the fringes from a number of spatially distinct identical triplets at a single instant.

The presence of instantaneously redundant baselines thus gives rise to a noise term very similar to that which is found in integrations which exceed the coherence time and which can be eliminated only by covering the aperture with a non-redundant aperture mask. This is because the point spread function of a non-redundant mask is unity, but the presence of redundant baselines destroys the coherence forcing the point spread function to depart from unity. It is clear from this analysis that the full aperture has some very undesirable properties which can be avoided by using masks, and this is therefore the preferred approach in observations with high photon rates. It is not quite so clear which approach is best in the case of low photon rates, but as this is not the subject of the present paper it will be discussed fully elsewhere.

In most situations ther are some baselines which are not matched one for one with baselines on the other two sides of the triangle. Suppose that there are n redundant triangles, and in addition there are p baselines ij, q baselines jk, and r baselines ki which are not matched on the other two sides. Then the triple product is:

$$\left(A_{ij_1}e^{i\phi_{ij_1}} + \ldots + A_{ij_n}e^{i\phi_{ij_n}} + A_{ij_{n+1}}e^{i\phi_{ij_{n+1}}} + \ldots \right.$$
$$\left. \ldots + A_{ij_{n+p}}e^{i\phi_{ij_{n+p}}}\right)\left(A_{jk_1}e^{i\phi_{jk_1}} + \ldots + A_{jk_n}e^{i\phi_{jk_n}} + A_{jk_{n+1}}e^{i\phi_{jk_{n+1}}}\right.$$
$$\left. + \ldots + A_{jk_{n+q}}e^{i\phi_{jk_{n+q}}}\right) \times \left(A_{ki_1}e^{i\phi_{ki_1}} + \ldots + A_{ki_n}e^{i\phi_{ki_n}} + \right.$$
$$\left. \ldots + A_{ki_{n+1}}e^{i\phi_{ki_{n+1}}} + \ldots + A_{ki_{n+r}}e^{i\phi_{ki_{n+r}}}\right).$$

Thus now, in addition to the terms of the first three kinds discussed above, there are terms of the kind

$$A_{ij_{n+s}} A_{jk_q} A_{ki_t} e^{i(\phi_{ij_{n+s}} + \phi_{jk_q} + \phi_{ki_t})}, \text{ etc. which we shall call}$$

'terms of the fourth kind'. There are $(p+q+r)n^2 + (pq+qr+rp)n + pqr$ terms of the fourth kind. The incoherent or noise term in the triple product thus has an rms amplitude of $[n^3+(p+q+r-3)n^2 + (pq+qr+rp+2)n + pqr]^{1/2}$, and in order to measure the closure phase to within an rms error of $1/x$ radians we require $N > x^2[n + (p+q+r-3) + (pq+qr+rp+2)/n + pqr/n^2]$.

The bi-spectrum approach works because the closure phase is preserved by the terms of the first kind. It is only because there are some closed triangles that any information about the phase is recoverable.

In the case of a filled aperture there are triangles for which $n \gg p$ and triangles for which $p \gg n$, etc., so the number of exposures required for a given signal-to-noise-ratio can be set by either of these.

We consider a square aperture and assume, for convenience, that the coherent patches are also square (see Figure 2), and we also assume, to begin with, that the phase distribution across the aperture can be approximated by a uniform grid of elementary areas of size r_o, each of which has a constant phase, with the value of the phase being selected randomly from the range $-\pi \rightarrow \pi$. This imposes much more regularity than the real situation, but leads to the correct qualitative results. The more realistic approach is most easily addressed by computer simulations, but some of the results which can be seen directly are discussed below. We also discuss, without loss of generality, only triangles for which two sides are parallel to the sides of the aperture. It is possible to cover the full complement of $2(m-1)^2$ independent closure phases in this way. In this case $r = 0$. Suppose that there are m^2 coherent patches over the whole aperture. The number of identical triangles is $n = (m-\ell_1+1)(m-\ell_2+1)$, where ℓ_1,ℓ_2 are the lengths of the sides of the triangle parallel to the sides of the aperture. We will discuss just two examples of triangles in such an aperture. These are marked "a" and "b" in Figure 2. Then we can calculate the quantities in Table I.

Thus, for example, for $m = 10$ the signal-to-noise-ratio for triangles of both types is 0.10. Note that we have only included the "noise" term due to the effect of redundant baselines on the closure phase, no terms due to photon noise, or other sources of error, have been included.

Experience with optical interferometry on the 5m telescope shows that we can use apertures of 20 cm diameter. In this case $m = 25$ and the signal-to-noise-ratio for triangles of both types is 0.040. Thus about 600 frames are needed to reduce the rms noise in the closure phase to below 1 radian. These errors are random and the closure phases are unbiassed so that good images can be made provided that enough closure phases are measured. There are $2(m-1)^2$ independent closure phases in an aperture consisting of m^2 elemental areas, so that in the case of the 5m telescope with 20 cm diameter coherence patches about 1000 independent closure phases can be measured.

TABLE I

	Triangles similar to "a"	Triangles similar to "b"
length of side ℓ_1	2	m
length of side ℓ_2	2	m
Number (= n) of similar triangles	$(m-1)^2$	1
number (= p) of unmatched, redundant baselines, ij.	m-1	m-1
number (= q) of unmatched, redundant baselines, jk.	m-1	m-1
number (= r) of unmatched, redundant baselines, ki.	0	0
Signal-to-noise-ratio $\left\{= \frac{1}{[n+p+q-3+(pq+2)/n]^{1/2}}\right\}$	$[(m-1)^2+2(m-1)+1-1/(m-1)^2]^{-1/2}$	$[(m-1)^2+2(m-1)]^{-1/2}$
SNR for large m	$\propto 1/m$	$\propto 1/m$

The signal-to-noise-ratio for large m is given approximately by $[(m-\ell_1)(m-\ell_2)+m^2/(m-\ell_1+1)(m-\ell_2+1)]^{-1/2}$. It is easy to show that this is approximately proportional to 1/m for most triangles. In fact only those with $\ell>0.9m$ deviate from this dependence by more than 10% for values of $m > 20$. There are very few of these, and so it is a good approximation to assume that the rms noise in the closure phases is proportional to m.

There are many ways of choosing the independent triangles. One could, for example, adopt a strategy in which the number of long baselines was minimized; or one in which the number of short baselines was minimised. The actual optimum strategy depends on the brightness distribution of the object under study. We have neglected the other effects, and these could make the signal-to-noise-ratio very small in some cases, for example on long baselines with heavily resolved objects.

We have seen that if m is large the rms closure phase error $\propto$ m, thus we can calculate the variation of sensitivity with m. We can think of the error in the closure phases arising from errors $\sqrt{3}$ times smaller on each baseline. We then have a situation similar to that in conventional aperture synthesis, in which the rms noise on the map is proportional to the rms noise on individual baselines, and inversely proportional to the total area (m^2) of the aperture. This shows that in the present example the noise will decrease in proportion to m, i.e. the sensitivity will increase in proportion to the <u>diameter</u> of the aperture, instead of the usual case in which it increases in proportion to the <u>area</u>. This is the sensitivity dependence that we get for incoherent, or post-detector, integration.

In the real situation we do not have a regular grid but a random distribution in two dimensions. In this case there are fewer identical triangles, and the probability, ε, of having triangles of a certain size and shape is a constant per unit aperture area which is determined by the characteristics of the scattering medium. The ratio of the coherent signal to the redundant baseline noise is:

$$\left(\varepsilon n + (p+q+r-3) + (pq+qr+rp+2)/(\varepsilon n) + pqr/(\varepsilon n)^2 \right)^{1/2} , \text{ for } \varepsilon n \geqslant 1$$

and 0 for $\varepsilon n < 1$. If $\varepsilon n \gg 1$ the signal-to-noie-ratio for large m is therefore still $\propto 1/m$ and we gain in sensitivity at most in proportion to m as the aperture size is increased. In practice it will often be the case that $\varepsilon n \sim 1$ so that the gain is even less.

3. CONCLUSION

The above analysis suggests that the advantage of fully filled apertures over non-redundant masks is not as great as might at first be supposed. For example, on the 5m telescope we may well be able to use a dozen 25 cm apertures. If these 12 apertures comprised a single aperture of equivalent size 87 cm, then the advantage of going to the full 5 m aperture would be less than a factor ten in sensitivity. In fact the non-redundant approach has <u>no</u> errors due to redundancy, so the gain will be even less. In addition, as we have seen, the irregularity of the phase variations across the aperture reduces the gains still further.

An alternative approach which should be explored would be to use masks with some redundancy and impose a 'minimum redundancy' condition which maximises the number of independent closure phases while restricting the number of redundant baselines. We know that we can compensate very well for non-uniform coverage of the aperture plane, and this is not exploited in the full aperture speckle approach.

4. REFERENCES

Baldwin, J.E., Haniff, C.A., Mackay, C.D., Warner, P.J. 1986. Nature 320, 595.

Cornwell, T.J., Wilkinson, P.N. 1981. MNRAS 196, 1067.

Cornwell, T.J. Astron. Astrophys. 1987 (in press).

Jennison, R.C. 1953. Ph.D. Thesis, University of Manchester.

Jennison, R.C. 1958. MNRAS 118, 276.

Lohmann, A.W., Weigelt, G. and Wirnitzer, B. 1983. Applied Optics. 24, 4028

Pearson, T.J., Readhead, A.C.S. 1984. ARAA 22, 97.

Readhead, A.C.S., Wilkinson, P.N. 1978. Ap.J. 223, 25.

Readhead, A.C.S., Walker, R.C., Pearson, T.J., Cohen, M.H. 1980. Nature 285, 137.

Readhead, A.C.S., Nakajima, T.S., Pearson, T.J., Neugebauer, G., Oke, J. B. and Sargent, W.L.W. 1987. Ap.J. (submitted)

Rogers, A.E.E., Hinteregger, H.F., Whitney, A.R., Counselman, C.C., Shapiro, I.I. et al. 1974. Ap.J. 193, 293.

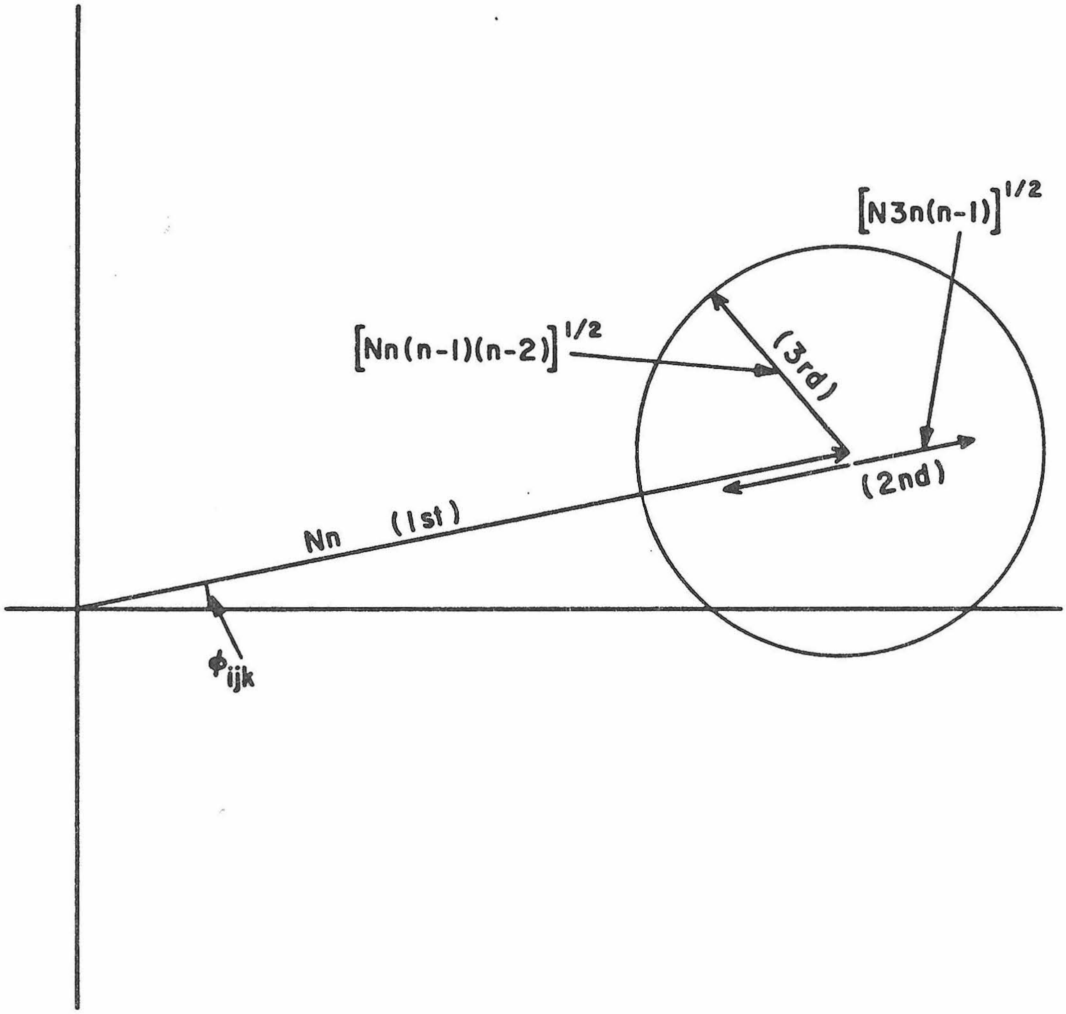
Schwab, F.R. 1980. Proc. Soc. Photo-Opt. Instrum. Eng. 231, 18.

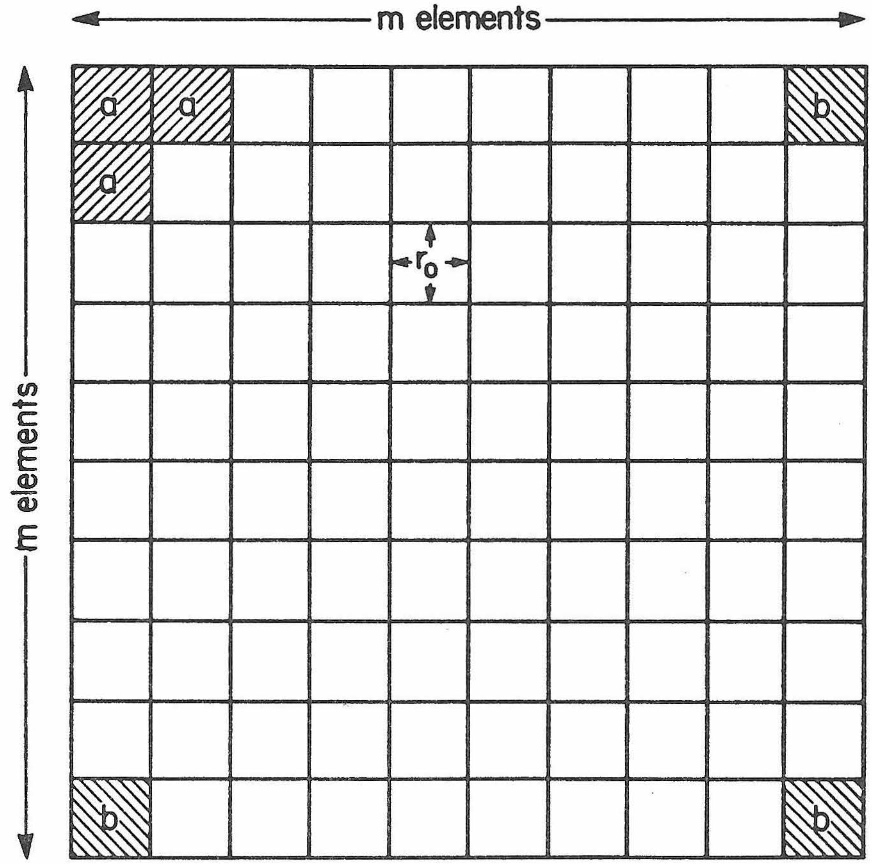
Weigelt, G., K.-H. Hofmann and Reinheimer, T. 1986. Proc. of the ESO Conf. "ESO's VEry Large Telescope II", Eds. D'Odorico, S. and Swings, J.-P.

5. FIGURE CAPTIONS

Figure 1. The effects of redundant baseline noise. The relative magnitudes of the contributions from terms of the first three kinds in the triple product are shown. Terms of the first kind add coherently. Terms of the second and third kinds add incoherently. In this example n = 10, i.e. the integration time is ten times the coherence time, or there are ten identical triplets of apertures contributing instantaneously to the triple product; and N = 100, i.e. the triple product has been averaged over 100 frames.

Figure 2. Coherent phase patches over a fully filled aperture. It is assumed that each elementary square, of area r_o^2 , has a constant phase, and that the phases change randomly between patches. This imposes more regularity than obtains in the real situation, but gives results which are qualitatively correct (see text).


[N3n(n-1)]^{1/2}
[Nn(n-1)(n-2)]^{1/2}
(3rd)
(2nd)
Nn (1st)
φijk


m elements
m elements
a
a
a
b
b
b
r_o