Radio Science, Volume 23, Number 4, Pages 713–720, July–August 1988

Satellite radar interferometry: Two-dimensional phase unwrapping

Richard M. Goldstein, Howard A. Zebker, and Charles L. Werner

Jet Propulsion Laboratory, California Institute of Technology, Pasadena

(Received October 13, 1987; revised March 23, 1988; accepted March 29, 1988.)

Interferometric synthetic aperture radar observations provide a means for obtaining high-resolution digital topographic maps from measurements of amplitude and phase of two complex radar images. The phase of the radar echoes may only be measured modulo 2π; however, the whole phase at each point in the image is needed to obtain elevations. We present here our approach to "unwrapping" the 2π ambiguities in the two-dimensional data set. We find that noise and geometrical radar layover corrupt our measurements locally, and these local errors can propagate to form global phase errors that affect the entire image. We show that the local errors, or residues, can be readily identified and avoided in the global phase estimation. We present a rectified digital topographic map derived from our unwrapped phase values.

INTRODUCTION

Satellite synthetic aperture radar interferometric techniques make use of phase measurements as well as the more conventional amplitude measurements. In this paper we describe a new phase estimation procedure that we have applied to the determination of ground topography from space. Once the phase differences have been obtained, we compute elevations by a method previously presented in detail [*Zebker and Goldstein*, 1986]. Here we focus on the problem of relating many individual phase measurements in a two-dimensional field by resolving the 2π ambiguities associated with the phase of signals.

A variety of signal processing problems require knowledge of relative phase between arbitrary points in one- or two-dimensional fields (see *Oppenheim and Lim* [1981] for a general review). A number of algorithms have been proposed to estimate phase from sampled, one-dimensional data (see, for example, *Tribolet* [1977]). Tribolet's algorithm and its successors are essentially adaptive schemes for integration of the phase derivative. *Hayes and Quatieri* [1983] propose a two-dimensional approach utilizing boundary value conditions. This type of algorithm is applicable to signals in which the magnitude of the Fourier transform is known and can be related through constraints to the amplitude and phase of the original signal.

Paper number 8S0268.
0048-6604/88/008S-0268$08.00

In this paper we examine in detail the two-dimensional phase estimation problem as it applies to radar interferometry. In this case we are not addressing the "phase retrieval" problem encountered in signal reconstruction, but instead we wish to determine the whole phase of signals associated with all points in an interferometric radar image where the amplitude and phase are essentially unrelated. We first outline the problem of estimating topography from radar signals and present the phase measurement problem. We then show that global errors in the estimated two-dimensional phases arise from "residues," local errors caused by noise in the signals or by actual discontinuities in the data. Next, we present an algorithm that identifies and isolates the residues so that a satisfactory set of phases is obtained. Finally, we present a topographic map derived from the radar interferometric data.

RADAR INTERFEROMETRIC TOPOGRAPHY

Radar interferometry [*Zebker and Goldstein*, 1986; *Zisk*, 1972*a*, *b*; *Rumsey et al.*, 1974; *Graham*, 1974] is a promising method of measuring topography which combines large coverage with high spatial resolution and good accuracy. For the parameters of our Seasat observations the spatial resolution was 50 by 50 m, and our elevation accuracy was better than 5 m over the flatter, brighter portions of the image. In this technique, two synthetic aperture radar images are acquired, not necessarily at the same time, with antennas displaced a certain distance (the interferometer baseline) across the line of motion. An inter-

ferogram is produced by averaging the corresponding amplitudes and differencing the corresponding phases at each point in the images.

Topography is directly related to the resultant phase of the interferogram. However, the phases are only measured modulo 2π. To calculate the elevation at each point in the image, the correct integer number of phase cycles must be added to each phase measurement; that is, the phase must be "unwrapped."

Plate 1 is an interferogram generated by combining two Seasat images of the Cottonball Basin of Death Valley acquired 3 days apart. Seasat orbited at an altitude of 800 km with an L band (23.5 cm wavelength) synthetic aperture radar aligned 22° up from the nadir. Because of the synchronicity of the Seasat orbits, the antenna separation baseline could be quite small; in this case it was 820 m. In the figure, brightness at each point represents echo power, and color represents phase. One complete cycle of phase corresponds to one revolution of a color wheel (cyan to magenta to yellow and back again to cyan). The phase measurements have been corrected for the expected phases of a topographically flat Earth, so the remaining phase is that due to topography; note the apparent relationship of this residual phase to altitude contours. The fringe pattern is associated with topography in the images and is used to form the final topographic map; hence accurate estimation of topography requires knowledge of the whole phase at each point. Since many cycles of phase are observed across the image, the fringes must be unwrapped.

GLOBAL ERRORS IN THE PHASE ESTIMATION

An obvious approach to phase estimation is to integrate phase differences from point to point, always adding the integer number of cycles that minimizes the phase differences. Consider the following one-dimensional sequence of phases:

$$0.5, 0.6, 0.7, 0.8, 0.9, 0.0, 0.1, 0.2, \cdots,$$

where the units are cycles. It is clear that one cycle should be added to the last three entries; the result is a phase ramp with no discontinuities.

Two types of errors are possible in the unwrapped sequence: (1) local errors, in which only a few points are corrupted by noise, and (2) global errors, in which the local error may be propagated down the entire sequence. We assume that the original scene is sampled often enough so that the true phase will not change by as much as one-half cycle per sample point. Layover, a departure from this assumption in which there is an actual discontinuity in the phase, will be discussed subsequently.

In the two-dimensional case we have a series of adjacent sequences of phase values. After integration of the phase differences in each line, global errors can occur where a substantial part of a sequence disagrees with its neighbor by more than one-half cycle, leading to a contradiction in the results of the phase unwrappings. No adjustment of integer numbers of cycles can remove the inconsistency.

We illustrate this inconsistency with the following two-dimensional field of "noisy" phase measurements:

0.0	0.1	0.2	0.3
0.0	0.0	0.3	0.4
0.9	0.8	0.6	0.5
0.8	0.8	0.7	0.6

We would like to reconstruct the whole phases from these measurements. Using the rule that no two adjacent points can differ by more than one-half cycle, we can fill in the additional cycles by scanning the points according to some rule, such as first scanning across each sequence and then down to align the sequences, resulting in the following solution:

0.0	0.1	0.2	0.3
0.0	0.0	0.3	0.4
−0.1	−0.2	−0.4	−0.5
−0.2	−0.2	−0.3	−0.4

Note that there are two values that differ from their vertical neighbors by more than one-half cycle. This is the contradiction described above that violates our assumption that no two points can differ by more than one-half cycle. The inconsistency in the phase resolution is apparent if we now choose another rule for scanning. If instead we scan down and then across, we obtain:

0.0	0.1	0.2	0.3
0.0	0.0	0.3	0.4
−0.1	−0.2	0.6	0.5
−0.2	−0.2	0.7	0.6

Not only are the estimated phases different in this case, but the location of the contradiction is changed.

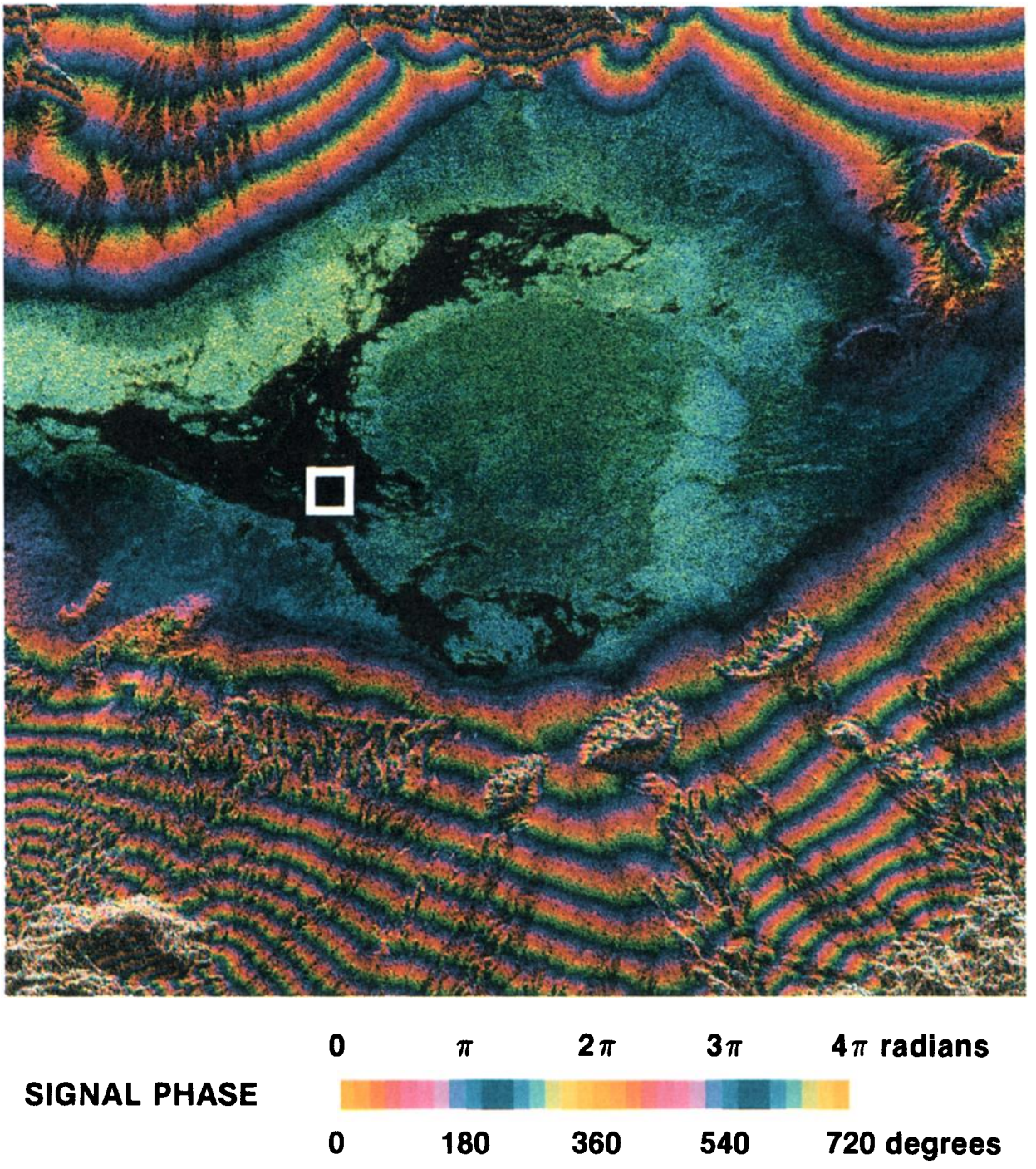

Plate 1. Interferogram generated by combining two Seasat images of the Cottonball Basin of Death Valley acquired 3 days apart. Brightness at each point represents echo power, and color represents phase. One complete cycle of phase corresponds to one revolution of a color wheel (cyan to magenta to yellow and back again to cyan). The phase measurements have been corrected for the expected phases of a topographically flat Earth, so the remaining phase is that due to topography; note the apparent relationship of this residual phase to altitude contours. Since many cycles of phase are observed across the image, the fringes must be unwrapped in order to compare topography at all points in the image.

This inconsistency is an inherent property of the data. No change in the order of scanning or in the allocation of integer cycles can eliminate it. Thus the two solutions are quite different when different paths of integration are chosen. Both solutions cannot be correct. The problem is then to determine the cause of the inconsistency and correct, or at least avoid it.

The reason for the inconsistency and a procedure for avoiding it become clear if we evaluate the sum of the phase differences clockwise around each set of four adjacent points: It is either zero, plus one cycle, or minus one cycle. Consider the following set of four points chosen from the center of the above example:

$$\begin{array}{ll} 0.0 & \rightarrow 0.3 \\ \uparrow & \quad \downarrow \\ 0.8 & \leftarrow 0.6 \end{array}$$

The net phase around the four points is plus one cycle, and no addition of cycles can eliminate it and still preserve the one-half cycle/point sampling criterion. If we evaluate these four-point integrations in the full data set, we obtain:

$$\begin{array}{cccc} 0.0 & 0.1 & 0.2 & 0.3 \\ & o & o & o \\ 0.0 & 0.0 & 0.3 & 0.4 \\ & o & +1 & o \\ 0.9 & 0.8 & 0.6 & 0.5 \\ & o & o & o \\ 0.8 & 0.8 & 0.7 & 0.6 \end{array}$$

We refer to these net resultant cycles as "residues" associated with the four points, where plus one cycle is a "plus" residue and minus one cycle is a "minus" residue. We can show that any integration path that encloses a single residue produces an inconsistency in the unwrapped phase and that the net phase difference integral around that path is nonzero. However, if a path encloses an equal number of plus and minus residues, no inconsistency results. Thus the phases can be unwrapped in a consistent manner if the residues are identified and suitable "branch cuts" are made between the residues to prevent any integration path from crossing these cuts. In the next section we show how the integration paths for phase unwrapping can be selected.

THE APPROACH

Filtering and residue reduction

As discussed above, residues arise from local errors in the measured phase caused by noise in the data and/or actual discontinuities caused by layover in the radar image. There are three major sources of noise in the phase measurements of an interferogram. First is receiver thermal noise. Radar design keeps this source low in comparison to echo power; however, there are always some dark areas in a scene which have low reflectivity, hence low signal-to-noise ratio.

Second is the "speckle" effect. Each resolution element in a radar image is generally composed of a collection of small scatterers. Each of these reflects with its individual amplitude and phase. We can easily show that the sum of these scattered signals is a random variable with variance equal to the square of the mean. Thus even for highly reflective areas a significant fraction of the image will have low signal-to-noise ratio.

Third is an antenna effect. Since the two antennas are not in exactly the same place, any resolution element appears slightly different in each image. The resulting phase noise depends on the antenna separation, the bandwidth of the radar, and the geometry of the imaging. For the Seasat case of Plate 1, separation was 20% of the critical separation which would completely decorrelate the two images. The critical baseline B is

$$B = \lambda R/(2 \text{ res})$$

where B is the baseline projected across the line of sight, res is the size of the resoluting element, similarly projected, λ is the wavelength, and R is the slant range.

The sum of the noises can be described by one dimensionless correlation coefficient, ρ. If $\rho = 1$, the phases in the two images are completely correlated and phase noise is zero; for $\rho = 0$, the phases of the interferogram are random and have no relation to the surface topography.

All of these noise sources can be mitigated by summing adjacent pixels of the interferogram presented above, taking account of both amplitude and phase. Of course, a loss of spatial resolution occurs. For the Seasat interferogram, four "looks" were summed and the resolution is 20 by 20 m. We have calculated, and show in Figure 1, the residue probability as a function of both the correlation coefficient and the number of looks. The calculations were done by simulation, using appropriate probability distributions for the noise and the signal; we note that the limiting value for $\rho = 0$ can be evaluated analytically and is equal to 1/3. Except near $\rho = 0$, one can see a dra-

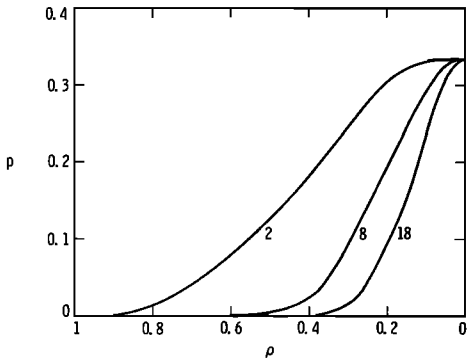

Fig. 1. Residue probability as a function of correlation coefficient and number of looks. Note that the probability drops as the number of looks increases; thus filtering can reduce residues at the expense of a loss of spatial resolution.

matic reduction in the number of residues expected as the number of looks increases. We have indicated points in Plate 1, in one of the dark regions, for testing. Filtering corresponding to 2, 8, and 18 looks was applied. The residue densities in these cases are 0.246, 0.051, and 0.015, respectively. These results imply that $\rho = 0.35$ for the dark regions of Plate 1 and give us confidence in our ability to predict the relationship of residues and noise.

Thus, filtering the signal will reduce the residues in any given image to make the selection of appropriate cuts feasible. The cost of this reduction will be a loss of spatial resolution in the final topographic map. We therefore reduce the number of residues in the interferogram by modest filtering. The filtering cannot be simple low pass, however, because each area will have a nonzero spatial fringe frequency owing to the geometry of imaging and the local slope of the surface. For Plate 1 the average fringe rate was 120 cycles per 1000 pixels. We compute a two-dimensional triangle function with a phase slope in the across-fringe direction that is equal to the average fringe rate and filter the entire image with that function. When applied to the interferogram of Plate 1, smoothing reduced the 77,500 residues to only 6,876 residues, and the spatial resolution is reduced to 50 by 50 m.

Branch cuts

What is needed next is to connect nearby plus and minus residues with cuts which interdict the integration paths, such that no net residues can be encircled and no global errors generated, although local errors in the immediate vicinity of the residues may occur. Of course, for pixels on opposite sides of a cut, there will be phase discontinuities of more than one-half cycle. Our goal, therefore, is to choose the cuts in such a way as to minimize the total length of the cuts and thereby minimize the total discontinuity.

Where the residue population is low, the location of optimum cuts is obvious. Where the density is high, however, one's ability to select cuts can be overwhelmed. For the noisier regions it seems that the only way to minimize the total length of the cuts is to try all $n(n-1)/2$ possibilities. Such a computationally intensive approach is not practical.

We have implemented the following algorithm to connect the residues with branch cuts. The interferogram is scanned until a residue is found. A box of size 3 is placed around the residue and is searched for another residue. If found, a cut is placed between them. If the residue is of opposite sign, the cut is designated "uncharged" and the scan continues for another residue. If, however, the sign of the residue is the same as the original, then the box is moved to the new residue and the search continues until either an opposite residue is located and the resulting total cut is uncharged or no new residues can be found within the boxes. In the latter case, the size of the box is increased by 2 and the algorithm repeats from the current starting residue.

The algorithm proceeds faster than the description of it. In the end, all of the residues lie on cuts which are uncharged, allowing no global errors. Where the residues are sparse, they are connected by cuts in an obvious way. Where they are very dense, whole areas are isolated. In effect, the algorithm "gives up." We have not obtained our goal of the unique, optimum solution. However, the approximation is excellent over most of the image, and, where it is not, the user is warned (by the density of the branch cuts).

We now discuss the particular type of residues associated with layover. The Seasat image coordinates are distance along track and slant range. Because the tops of mountains and hills can be closer to the radar than their bases, they appear in the image to lean toward the radar. This is the layover effect, and it is pronounced in Seasat images because the radar beam was pointed only 22° away from the nadir. Where layover occurs, there is a true discontinuity in the interferogram, one which is not caused by noise. These areas are typically characterized by a preponderance of residues of similar charge in a line on one half of the layover region and a corresponding set of

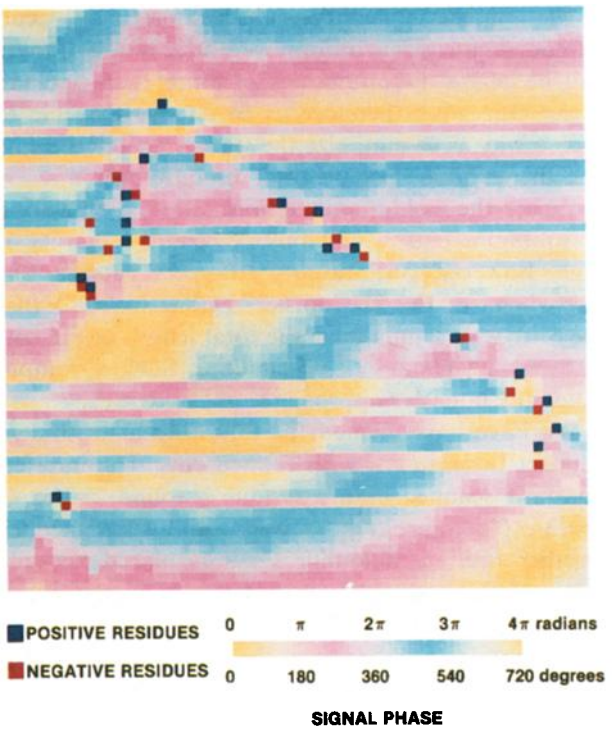

Plate 2*a*.

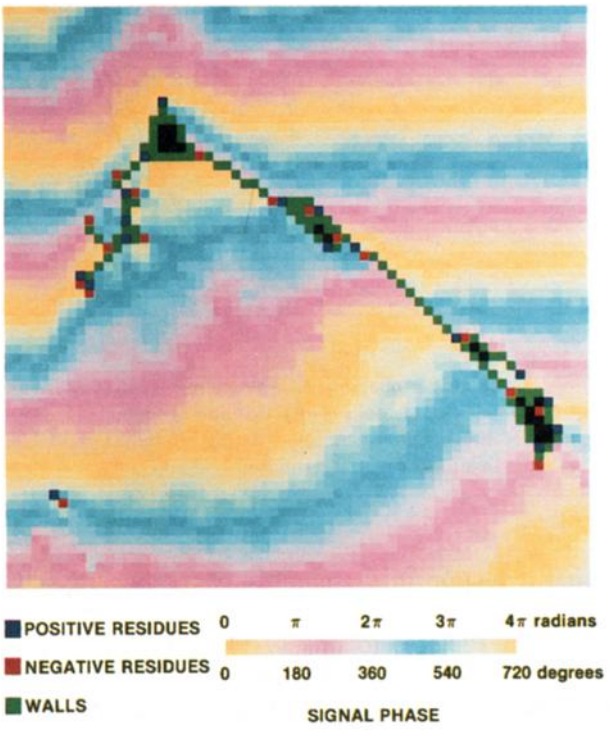

Plate 2*c*.

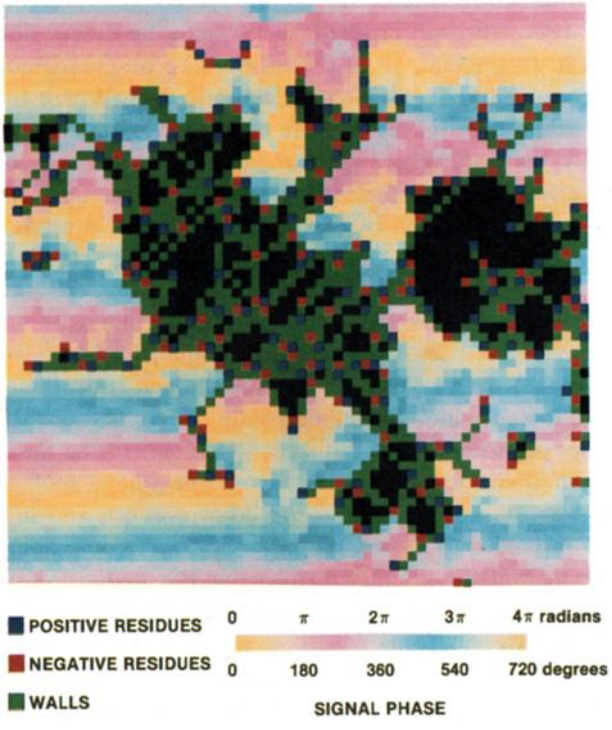

Plate 2*b*.

Plate 2*a*. Enlarged portion of Plate 1 with phases blindly unwrapped by choosing closest 2π multiple. In this plate, two cycles in phase are represented by one revolution of the color wheel; therefore a one-cycle error will show up with maximum clarity as a shift halfway around the color wheel. A global error (discontinuity) radiates from every residue.

Plate 2*b*. Same region as Plate 2*a*, but with cuts in place before unwrapping. The laid over part in the center is now isolated properly to prevent global errors.

Plate 2*c*. Another region from Plate 1, chosen to illustrate the unwrapping results in a residue-dense region; this area is entirely isolated from phase estimation. This is a case where the algorithm "gives up," as no reliable estimate of phase is possible.

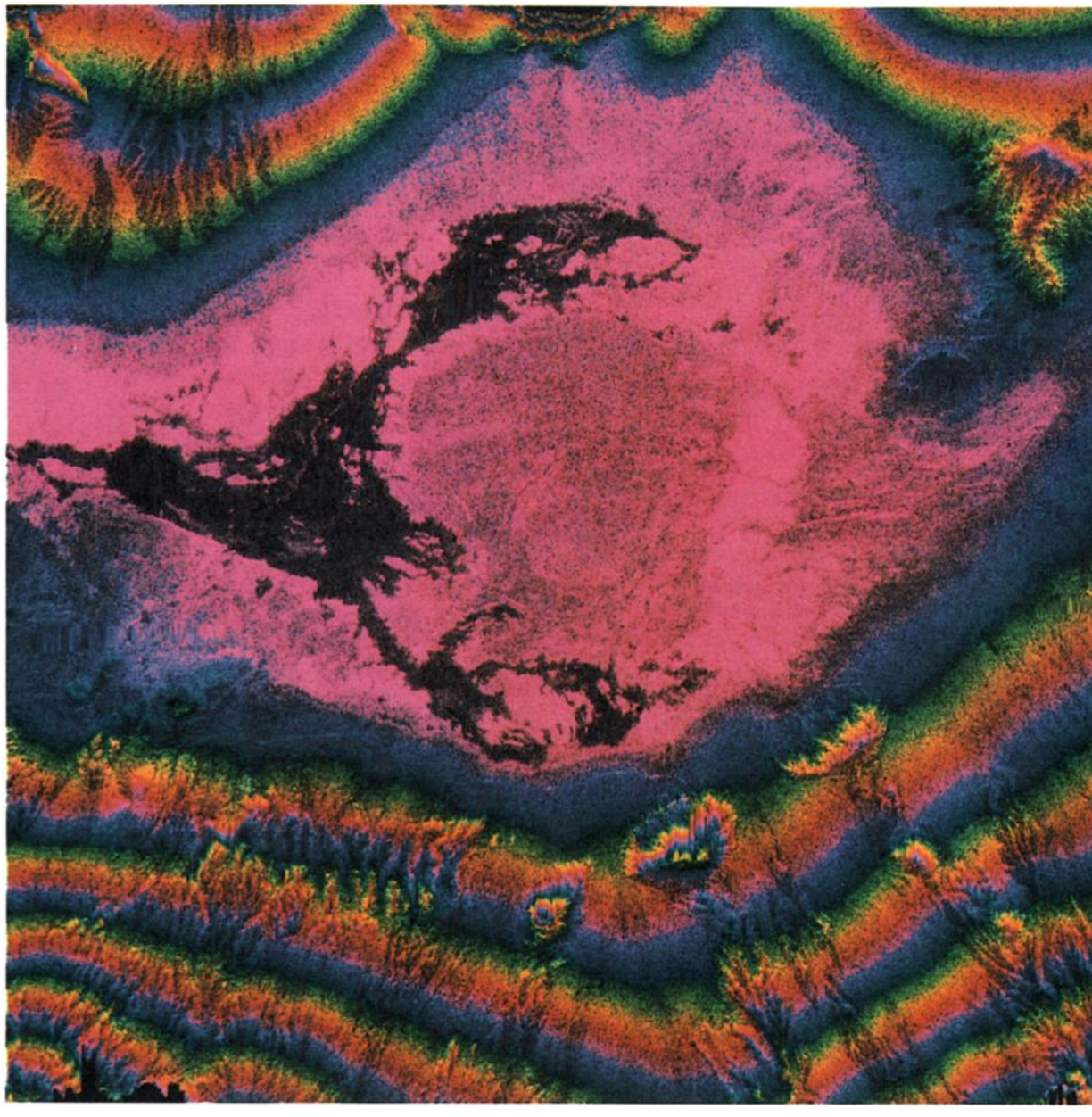

Plate 3. Final topographic map of region in Plate 1 derived from unwrapped phase values. This map has been rectified to ground range and along-track coordinates. This map is no longer laid over as are typical radar images, and the contours are in close agreement with published U.S. Geological Survey maps.

opposite charges on the other half (see Plate 2*a*). Our cutting algorithm is successful in isolating such areas.

Integration and examples

The final step is to integrate the phase differences in such a way as not to cross any of the cuts. We use a routine wherein at each point for which we have obtained a phase estimate, a new estimate is applied to any neighbor that does not already have an estimate and that is not on the other side of a cut. At each application of the algorithm the appropriate integer number of cycles is added to the phase such that the phase difference is always less than one-half cycle. Our data are quantized to an odd number of levels so that a phase difference of exactly one-half cycle never occurs.

Plate 2*a* shows an enlargement of the marked portion of Plate 1 where a phase estimate was obtained before any cuts were made. In this plate two cycles in phase are represented by one revolution of the color wheel; therefore a one-cycle error will show up with maximum clarity as a shift halfway around the color wheel. A global error (discontinuity) radiates from every residue. Plate 2*b* shows the same region but with cuts in place before the integration is performed. The laid over part in the center is isolated properly to prevent global errors. In Plate 2*c* we show another region from Plate 1, chosen to illustrate the results in a residue-dense region; this area is entirely isolated from phase estimation. This is a case where the algorithm "gives up," as no reliable estimate of phase is possible.

THE RESULTING TOPOGRAPHIC MAP

Finally, we show in Plate 3 the results of unwrapping the image in Plate 1, calculating the altitude contours using the method of *Zebker and Goldstein* [1986] and rectifying the image to ground range and along-track coordinates. This map is no longer laid over as are typical radar images, and the contours are in close agreement with published U.S. Geological Survey (USGS) maps. We are presently devising techniques for comparison of our digital terrain map with the digital models produced by the USGS by digitizing existing contour maps. Lack of high-resolution digital maps precludes a detailed error analysis, but it illustrates the need for efficient topographic mapping procedures that can present results in digital form. The satellite interferometric approach yields this product directly, without time-consuming hand digitization of contour maps obtained by conventional stereogrammetric techniques.

Acknowledgments. This work was supported in part by the Land Processes Branch of the National Aeronautics and Space Administration.

REFERENCES

Graham, L. C., Synthetic interferometer radar for topographic mapping, *Proc. IEEE, 62*, 763–768, 1974.

Hayes, M. H., and T. F. Quatieri, Recursive phase retrieval using boundary conditions, *J. Opt. Soc. Am., 73*, 1427–1433, 1983.

Oppenheim, A. V., and J. S. Lim, The importance of phase in signals, *Proc. IEEE, 69*, 529–541, 1981.

Rumsey, H. C., G. A. Morris, R. R. Green, and R. M. Goldstein, A radar brightness and altitude image of a portion of Venus, *Icarus, 23*, 1–7, 1974.

Tribolet, J. M., A new phase unwrapping algorithm, *IEEE Trans. Acoust. Speech Signal Process., 25*, 170–177, 1977.

Zebker, H. A., and R. M. Goldstein, Topographic mapping from interferometric synthetic aperture radar observations, *J. Geophys. Res., 91*, 4993–4999, 1986.

Zisk, S. H., A new, Earth-based radar technique for the measurement of lunar topography, *Moon, 4*, 296–300, 1972*a*.

Zisk, S. H., Lunar topography: First radar-interferometer measurements of the Alphonsus-Ptolemaeus-Arzachel region, *Science, 178*, 977–980, 1972*b*.

R. M. Goldstein, C. L. Werner, and H. A. Zebker, Mail Stop 300-235, Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109.