Three- and Four-Satellite Continuous-Coverage Constellations

John E. Draim*
ANSER (Analytic Services Inc.), Arlington, Virginia

This paper addresses the problem of obtaining continuous satellite line-of-sight coverage of every point in either the northern hemisphere or on the entire Earth's surface with the minimum number of satellites. Geometric theorems and corollaries relating to coverage are presented. A three-satellite elliptic-orbit constellation covering the entire northern hemisphere is described. Also, a four-satellite constellation giving continuous global coverage is defined. An extensive search has uncovered no other three-satellite continuous hemispheric or four-satellite continuous global coverage constellations.

Nomenclature

e	= eccentricity
i	= inclination
M	= mean anomaly
p	= perpendicular distance from center of Earth to satellite plane
R_E	= mean radius of Earth (= 3442 n.mi.)
$S_1,S_2,S_3,S_4...$	= satellite designators
$S_1',S_2',S_3',S_4'...$	= satellite suborbital points
T_c or $T_{\rm const}$	= constellation period, h (equals elliptic orbit period)
σ	= satellite visibility angle ("look angle") above horizon
Ω_c	= constellation frequency
Ω_{op}	= osculating or rotating plane frequency

Background

THE question "How many satellites are required to achieve continuous global coverage?" has been discussed by a number of authors. In 1969, Roger Easton of the Naval Research Laboratory stated that "a minimum of six satellites is required in a controlled-orbit constellation to have at least one satellite continuously visible from all points on the earth's surface."[1] Versions of Easton's constellation are shown in Fig. 1. In 1970, John G. Walker of England's Royal Aircraft Establishment stated that "whole earth coverage cannot be maintained at all times with less than five satellites."[2] Walker's constellation is shown in Fig. 2. In 1980, A. H. Ballard of TRW wrote that "a total of at least five satellites is required for worldwide single visibility."[3] All three authors have assumed that the constellations employ circular, common-period orbits, although only Walker and Ballard explicitly state this assumption. Walker also stated that "while elliptical orbits have some advantages in the provision of coverage to limited areas mainly confined to either the northern or the southern hemisphere, circular orbits appear to have the advantage for zones extending equally into both hemispheres, and even more as regards whole earth coverage."

The author's goal has been to reduce, if possible, the number of satellites required for continuous global or hemispheric coverage. Reduction in the number of satellites should lead to significant booster and satellites cost savings. Alternatively, improvements in coverage should be possible, using the same number of satellites.

Introduction

The purpose of this paper is to incorporate the use of elliptic orbits into the design of multisatellite constellations. It will be demonstrated that the general assumption that circular orbits have the advantage as regards whole-Earth coverage is in error. The determination that at least five satellites are required to provide whole-Earth coverage, while true for circular orbits, does not apply to elliptic-orbit constellations. In fact, it will be shown that continuous global coverage can be achieved with only four elliptic-orbit satellites and that continuous hemispheric coverage can be

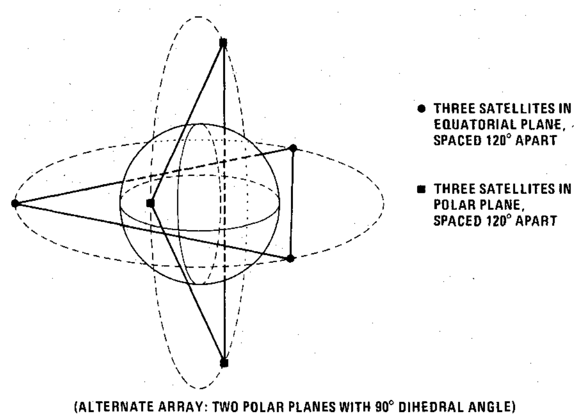


Fig. 1 Six-satellite continuous global coverage model proposed by Roger Easton, 1969.

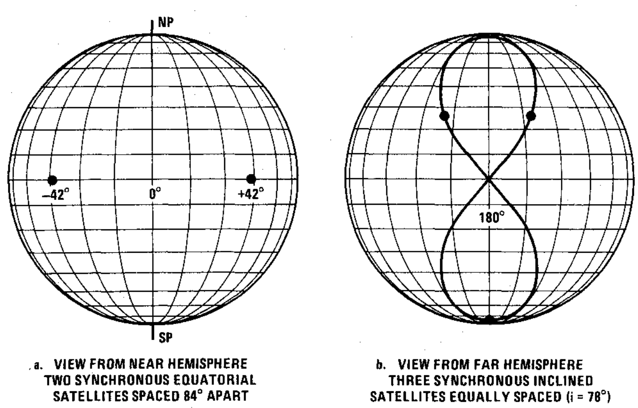


Fig. 2 Five-satellite continuous global coverage model proposed by John Walker, 1970.

Presented as Paper 84-1996 at the AIAA/AAS Astrodynamics Conference, Seattle, WA, Aug. 20-22, 1984; submitted Oct. 9, 1984; revision received March 18, 1985.

*Principal Aerospace Engineer, Space Systems Division; currently Director for Advanced Concepts, Science and Technology Associates, Inc. Member AIAA.

achieved with only three elliptic-orbit satellites. The additional flexibility obtained through the use of elliptic orbits opens up new possibilities in constellation design.

For the restricted case of circular orbits, it can be shown that three satellites in noncoplanar orbits must at some point in time lie along the arc of a great circle (i.e., the satellite plane must at some point in time contain the Earth's center). The two resultant polar coverage gaps (lying at the extremities of the Earth diameter normal to the satellite plane) thus make continuous hemispheric coverage impossible with only three circular orbit satellites. Since at least one additional satellite is needed to cover up one of these polar gaps, a minimum of four satellites in circular orbits is required for continuous hemispheric coverage. Similarly, since at least two extra satellites are required to cover both polar gaps, complete and continuous global coverage requires a minimum of five circular-orbit satellites.

The employment of circular synchronous equatorial orbits results in coverage gaps at and near the poles of the Earth, regardless of how many satellites are used. For example, three stationary satellites spaced 120 deg apart on the equator leave a triangular gap at both the North and South Poles. For certain mission requirements, it is desirable to close one or the other, or both, of these gaps. Generally, the northern hemisphere is of greater interest and coverage of the North Pole and the entire northern hemisphere is given priority. A common solution consists of adding an extra ring of satellites in inclined orbits. This practice can be quite expensive given the high costs of producing, launching, and maintaining these extra satellites on orbit. Continuous hemispherical or global coverage with the minimum number of satellites thus has obvious advantages.

Coverage Theorems

As a first step, the sufficient conditions for complete, instantaneous coverage will be stated in the form of geometric theorems. For reasons of brevity, proofs of these theorems will not be given. Later, the conditions for continuous coverage will be addressed. For simplicity, a spherical Earth is assumed.

Theorem I states the sufficient conditions to ensure complete, instantaneous coverage at all points within the spherical triangle formed by the suborbital points of three satellites (see Fig. 3).

Theorem I: If a plane containing three satellites does not intersect the Earth or if this plane is tangent to the Earth at a point, then every point within the spherical triangle on the Earth's surface formed by the satellites' suborbital points is visible from at least one of the satellites.

Theorem II states the sufficient conditions for visibility within specified Earth-surface areas defined by two satellite suborbital points and a great circle (see Fig. 4). The areas in question may be either two spherical triangles (case 1) or a spherical quadrilateral (case 2).

Theorem II: If a plane containing two satellites is perpendicular to a great circle on Earth and, if this plane either does not intersect the Earth or is tangent to the Earth (at a single point on the great circle), then all points within the area (or areas) encompassed by the great circle, the smaller arc between the two subsatellite points, and the arcs from each subsatellite point perpendicular to the great circle are visible from at least one of the two satellites.

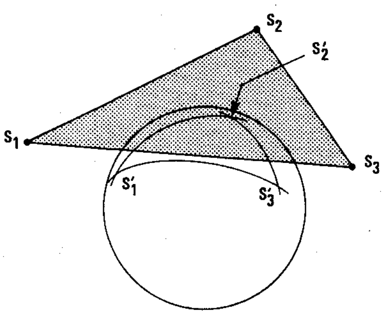


Fig. 3 Theorem I.

In order to apply the two theorems just stated to ensure instantaneous northern hemisphere coverage, one constructs a prism using the three satellites, S_1, S_2, and S_3. The top of the prism is formed by the plane that passes through all three satellites. The sides of the prism are formed by selecting two satellites at a time and passing planes through them that are perpendicular to the plane of the equator. It will also be assumed that the prism sides (the planes perpendicular to the equator) are infinitely long in the downward direction, so that there is no "bottom" to the prism. Then, applying theorem I to the top plane and theorem II three times in succession to the three side planes, it is evident that total instantaneous coverage of the entire northern hemisphere is obtained, provided that the prism always surrounds, but does not intersect, the Earth. (Coverage will again be assumed if a plane is merely tangent at a single point on the sphere.) This construction is covered in corollary I (see Fig. 5):

Corollary I: A four-sided prism is formed by passing a plane through all three satellites, and then three planes through pairs of satellites in sequence that are perpendicular to a given great circle. If none of the planes of this prism intersects the Earth, then all points within the hemisphere bounded by the great circle and lying on the side of the three-satellite plane are visible from at least one of the three satellites, by theorems I and II.

To provide for continuous hemispheric coverage in time, it is necessary only to apply the theorems given above repetitively to the irregular prism as it warps continuously through an entire constellation period. If, as shown in Fig. 5, the prism always surrounds the Earth without intersecting it, then continuous and complete coverage of the northern hemisphere is ensured.

To provide for instantaneous complete global coverage by four satellites, in a similar manner, it is necessary to apply

Fig. 4 Theorem II.

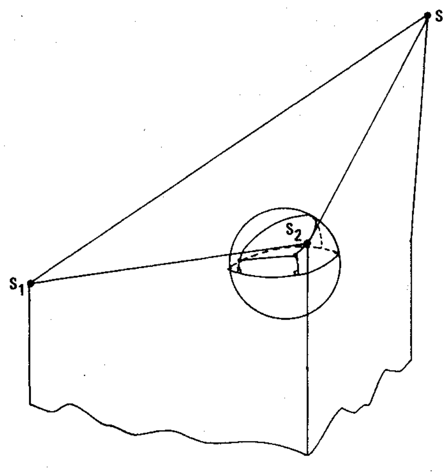


Fig. 5 Corollary I.

theorem I and corollary II (Fig. 6) four times successively. The four spherical triangles defined by the suborbital points will cover the entire surface of the Earth. If the tetrahedron surrounds the Earth without contacting it or is at the most tangent at a point or points, then instantaneous global coverage will result. That is, at least one satellite will be visible from any point of the Earth's surface at that instant.

Corollary II: Pass planes through three satellites at a time in a four-satellite constellation; the subsatellite points taken three at a time define four spherical triangles. It can be shown that, if none of the four planes intersect the Earth's surface and if the Earth is completely enclosed within the tetrahedron formed by the four planes, then any point on the Earth's surface is visible from at least one of the satellites, by using theorem I.

To obtain complete, continuous global coverage, it is necessary to show that the planes of the tetrahedron always encompass the Earth, without intersecting it, as the tetrahedron changes shape or warps during the constellation period. The planes can momentarily become tangent to the Earth's surface, but the visibility criteria are still met. This leads to the use of the descriptive term "osculating planes." The tangent condition will occur for some critical minimum constellation period, where the planes momentarily "kiss" the Earth's surface and then rise to a higher altitude.

Computer models were developed to calculate the perpendicular distances from the center of the Earth to any of the above planes. If the perpendicular distance is less than the Earth's radius, then the plane will intersect the Earth. If, on the other hand, the perpendicular distances are always greater than the Earth's radius, then it is evident that the plane in question does not intersect the Earth's surface.

Since we are concerned with coverage of either the entire globe or the entire northern hemisphere, it does not matter whether we introduce Earth rotation or not. As a matter of convenience, the orbital equations have been written in inertial space. Also, orbital perturbations due to Earth oblateness and lunar attraction have not been considered here. Their magnitudes are small and in practice could be cancelled by small periodic orbital corrections.

The Basic (Circular) Cubic Constellation

Satellite constellations have been constructed with the orbital planes lying parallel to the faces of a regular polyhedron.[4] Research by the present author has shown ways of using other polytopes (solid polyhedra or plane polygons) to space the satellites within these orbital planes. In the case of a three-satellite array, three faces of a cube form a convenient starting point. If a cube is tipped so that a major diagonal is vertical, then each plane face of the cube will be inclined at 54.735 deg from the horizontal. Each of the planes will be spaced 120 deg apart on the equator. For each satellite, a line of nodes is established (which will be used for eventual rotation into other inclinations). For each orbit, an argument of perigee of −90 deg is specified (to place each perigee in the southern hemisphere). This provides a reference for measuring either true or mean anomaly and allows for the later perturbation from circular to elliptic orbits. For the starting position, we place the satellites into circular orbits at true anomalies of 0, 120, and 240 deg, respectively, progressing to the east in the inertial reference frame. That is, satellite S_1 is placed at a true anomaly of 0 deg, in an inertial frame, with its line of nodes at the reference value of 0 deg. Satellite S_2, with a line of nodes located at 120 deg to the east of satellite S_1, is given a true anomaly of 120 deg. (This displaces satellite S_2 roughly 240 deg to the east of S_1.) Finally, satellite S_3, whose line of nodes is 240 deg to the east of S_1's, is assigned a true anomaly of 240 deg. This puts it roughly back at 120 deg to the east of S_1. From the starting positions, circular orbital motion is then allowed to proceed through one complete constellation orbital period (see Fig. 7). A constellation orbital period is defined as equal to an individual satellite period.

Examining the characteristics of the satellite plane of the basic circular-orbit cubic constellation throughout a complete constellation period, some interesting phenomena are noted. First, the perpendicular to the satellite plane remains at a fixed angle from the polar axis and rotates around it. This angle is found to be 45.99 deg. Second, the perpendicular to the satellite plane rotates around the polar axis at twice the constellation frequency (see Fig. 8). Third, the satellite plane does not continuously pass through the center of the Earth. Instead, it moves up and down on the polar axis, with a frequency of motion three times the constellation frequency (see Fig. 9). Thus, six times per constellation

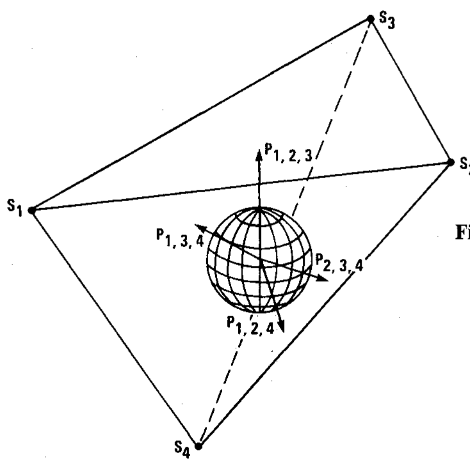


Fig. 6 Corollary II.

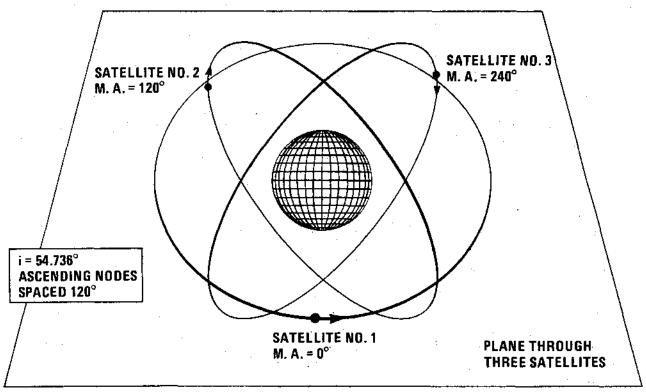


Fig. 7 Basic cubic constellation showing rotating plane at $t=0$ (circular orbits).

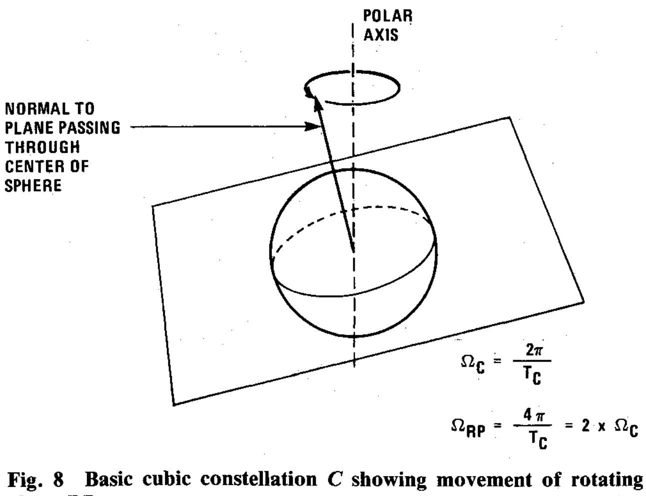


Fig. 8 Basic cubic constellation C showing movement of rotating plane RP.

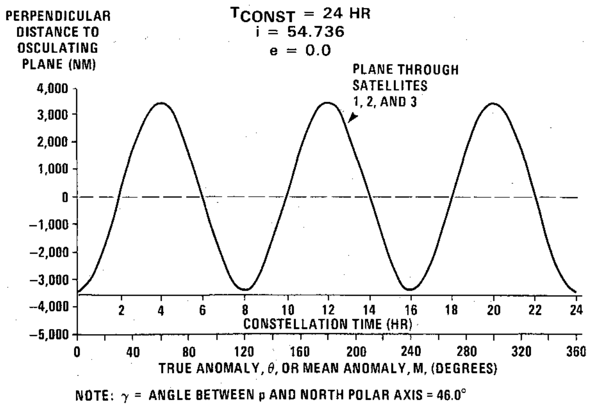


Fig. 9 Perpendicular distance p to satellite plane from Earth center, circular cubic constellation.

period, the satellite plane passes through the center of the Earth. Coverage gaps will be offset from the polar axis due to the tilt of the plane.

A computer model was developed that presents the approximate visibility arcs for any set of satellites at any instant during the constellation period, looking down from the North Pole. For this circular-orbit constellation, the picture looking up from the South Pole would be identical, with a phase difference. Figure 10 shows how the gaps rotate around either polar axis at twice the constellation frequency. The southern hemisphere coverage would be the same as the northern, with a phase difference.

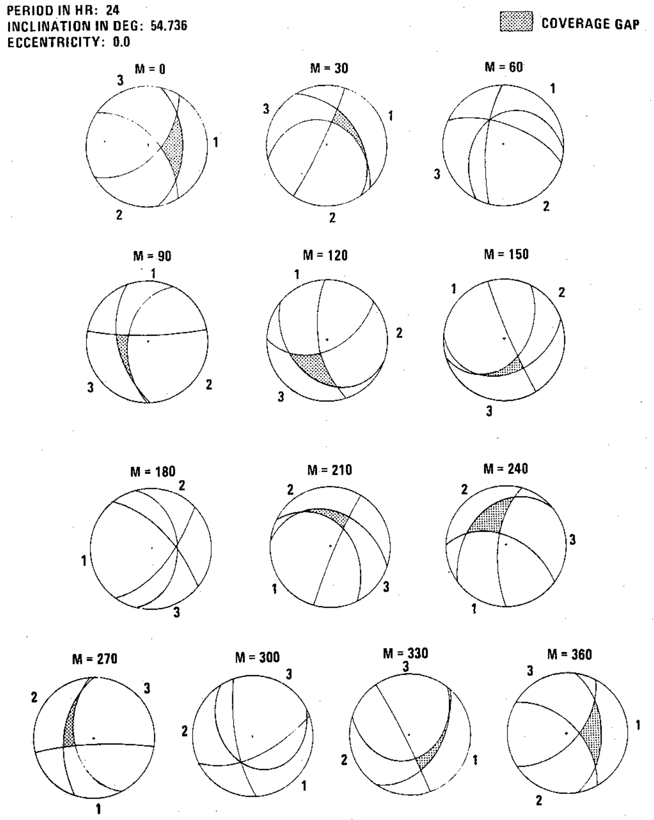


Fig. 10 Basic circular cubic three-satellite constellation, northern hemisphere coverage arcs, looking down from North Pole.

The Three-Satellite Elliptic Cubic Constellation for Continuous Northern Hemisphere Coverage†

If the basic (circular-orbit) cubic constellation just described is now altered by reducing the inclination of all three planes, introducing an eccentricity, and substituting mean anomaly for true anomaly at all points in the orbits (to accommodate the change from circular to elliptic orbits), the northern hemisphere coverage gap can be suppressed throughout the constellation period. This is accomplished at the expense of allowing the size of the southern hemisphere gap to increase. Proper selection of the inclination-eccentricity combinations will lead to an optimal solution marked by the maximization of the minimum value for the perpendicular distance p.

For this elliptic-orbit example, 24 h was used for the period, 30 deg for the orbital inclination, and 0.28 for the eccentricity of all three orbits. An isometric diagram of the perturbed, elliptic cubic constellation is shown in Fig. 11. In effect, we have raised the satellite plane above the northern hemisphere for an entire orbital period. The plane can be made to osculate (make contact momentarily) or to maintain some minimum separation distance from the Earth's surface. The minimum-visibility look angle (angle at which satellites are viewed above the horizon) will be a function of this separation distance.

To see the effect of this synchronous (24 h) constellation in terms of coverage and ground tracks, both the continuous coverage areas and satellite ground tracks have been plotted on a standard Mercator chart (see Fig. 12). Since there is no restriction in selecting the longitudes of the ascending nodes (so long as they are kept 120 deg apart), values are selected that provide coverage to the major southern hemisphere land masses of Africa, Australia, and South America. It can be seen in Fig. 12 that continuous coverage is thus provided to most of the world's populated land masses with just three

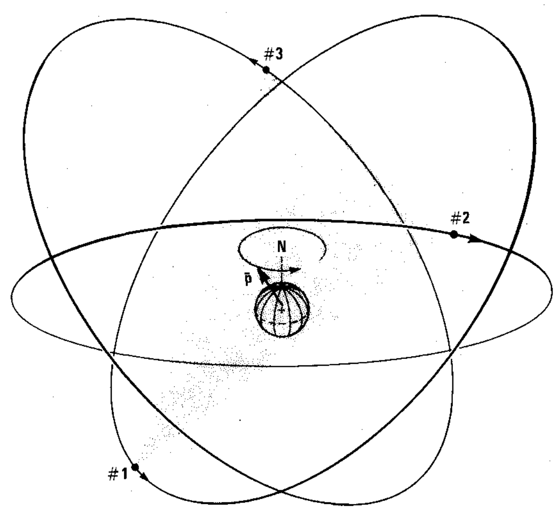


Fig. 11 Elliptic cubic three-satellite constellation giving continuous coverage of entire northern hemisphere.

satellites. It should also be noted that there is a minimum orbital altitude for continuous hemispherical coverage with this cubic constellation. More precisely, a particular value of semimajor axis, corresponding to a unique period, will define this minimum; any lower values will necessarily lead to an intersection of the Earth's surface by the satellite plane or planes. For the northern hemisphere, continuous-coverage constellation, this minimum constellation period was determined to be 16.1 h. For constellation periods greater than this minimum, a range of inclination/eccentricity combinations can be used. However, it should be noted that the southern hemisphere land masses can be provided continuous coverage only with a 24-h (synchronous) constellation period.

†U.S. patent pending.

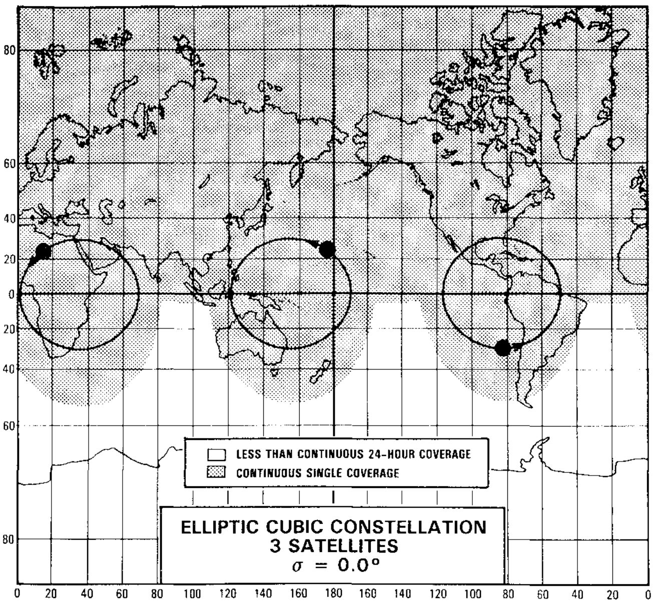


Fig. 12 Geographical coverage and ground tracks of elliptic cubic constellation.

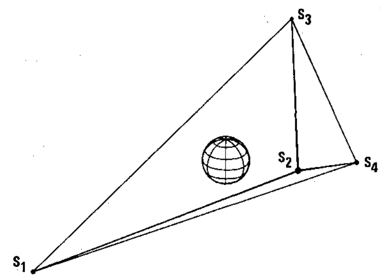
The Four-Satellite Continuous Global Coverage Constellation‡

The underlying concept of covering the globe continuously with four satellites involves exploiting the facts that 1) the only remaining gap in coverage occurs in the southern hemisphere and 2) this gap rotates at twice the constellation frequency. The reasoning is that a single circular-orbit satellite (in addition to the original three), having a frequency of rotation twice that of the other three and in an equatorial orbit, can cover the remaining southern hemisphere gap. However, for this to happen, it is evident that the original three satellites have not only to provide continuous coverage of the entire northern hemisphere, but also to cover the South Pole and the area immediately surrounding it. It was found that this could be accomplished, providing the altitude is high enough (period lengthy enough) and the eccentricity does not exceed a critical value. A 96 h constellation (elliptic-orbit) period will meet these stringent additional requirements. This period was selected because it is an integral multiple of 24 h and also because it provides some "cushion" in keeping the osculating tetrahedral planes from actually contacting the Earth's surface. Since the basic period of the three original satellites is 96 h, this means that the circular equatorial satellite has a period of 48 h. The satellite parameters, along with an isometric view of the constellation, are shown in Fig. 13.

Using the visibility arc computer model, snapshots of coverage arcs for every 30 deg of constellation mean anomaly are shown from the South Pole aspect. Only the three common-period satellites are shown, so that the remaining gap to be covered by the fourth satellite will be visible. There are no coverage gaps in the northern hemisphere throughout the constellation period. Looking up from the Southern Hemisphere, it can be seen that there is a continuous coverage gap, which does, indeed, rotate at twice the constellation frequency around the South Pole. It can also be seen that the original three satellites do continuously cover the South Pole itself, plus a region of a few degrees on all sides of this pole (see Fig. 14). The remaining coverage gap will now be covered by the circular equatorial 48-h period satellite.

As a final check, the perpendicular distances from the center of the Earth to all four planes of the constantly changing tetrahedron were calculated for small increments of mean anomaly during a complete constellation period (see Fig. 15). It can be seen that the closest approaches of all of the planes occur at roughly 4000 n.mi. from the Earth's center. Referring back to theorem I and corollary II, the Earth is continuously covered by the four satellites for a complete constellation period (and thus for all succeeding periods). It should be noted that the plane containing the three 96-h satellites approaches the Earth at minimum distance three times per orbital period, while the other three planes (each of which contains the 48-h satellite) approach to a minimum distance only once per orbital period. Also, the maximum separations of the latter three planes is much greater, which led to plotting the distances logarithmically. The critical constellation period for obtaining continuous global coverage is

‡U.S. patent pending.



	PERIOD (HOURS)	ECCENTRICITY	INCLINATION (DEGREES)	RIGHT ASCENSION OF ASCENDING NODE* (DEGREES)	ARGUMENT OF PERIGEE (DEGREES)	MEAN ANOMALY (DEGREES)
SATELLITE 1	96	0.13	33	0	−90	0
SATELLITE 2	96	0.13	33	120	−90	120
SATELLITE 3	96	0.13	33	240	−90	240
SATELLITE 4	48	0.00	0	180	0	0

*INERTIAL REFERENCE FRAME

Fig. 13 Four-satellite global coverage constellation.

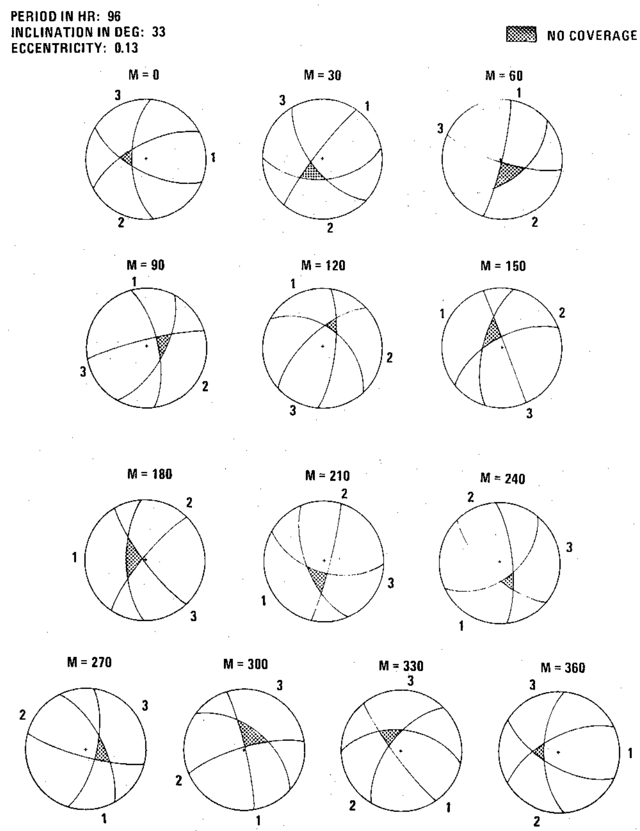


Fig. 14 Four-satellite global coverage model, southern hemisphere coverage arcs for satellites 1-3, looking up from the South Pole (satellite 4 omitted).

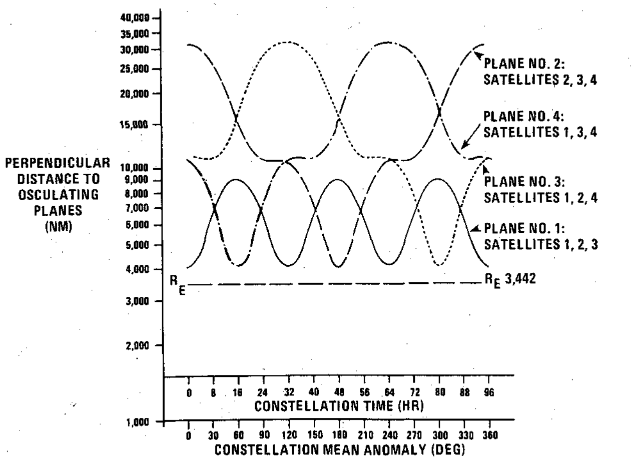


Fig. 15 Perpendicular distance from Earth center to the four osculating planes in a four-satellite global coverage model ($T_C = 96$ h).

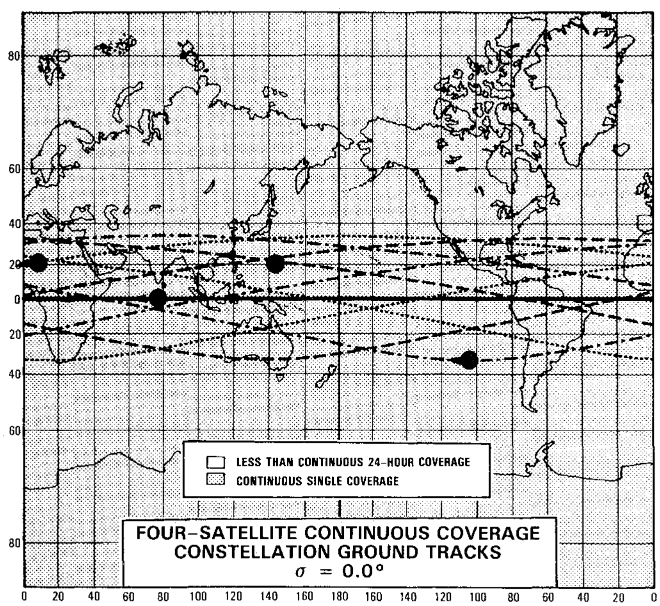


Fig. 16 Geographical coverage and satellite ground tracks of four-satellite continuous global coverage array.

78 h, which means that the circular equatorial satellite must have a period of 39 h.

A Mercator coverage and satellite ground track chart is shown in Fig. 16 for the 96-h constellation. Since the three "top" satellites are in inclined orbits with a 4-day period, they appear to rotate from east to west in the sky once every 72 h. The circular-orbit equatorial satellite, having a 2-day period, also rotates from east to west but with an apparent 24-h period. Since the orbits are fairly low in inclination, there is somewhat more redundant coverage in the equatorial regions than in the polar regions; but there is complete coverage by at least one satellite continuously, everywhere on Earth.

Conclusions

The generally accepted hypothesis that circular-orbit constellations are superior to elliptic-orbit constellations in providing continuous global coverage with the minimum number of satellites is shown to be in error. Specifically, it is shown in this paper that:

1) Using circular orbits, continuous hemispherical Earth coverage cannot be provided by three satellites and continuous global Earth coverage cannot be provided with four satellites.

2) Continuous northern (or southern) hemispherical coverage can be provided with a three-satellite constellation using inclined elliptic orbits for constellation periods of 16.1 h or greater.

3) Continuous global Earth coverage can be provided with a four-satellite constellation having three inclined elliptic orbit satellites and a single circular equatorial satellite having one-half the period of the rest of the constellation, for constellation periods of 78 h or greater.

Utilizing the principles embodied in these concepts for multisatellite constellation design, it is possible that significant improvements in coverage and/or cost reduction may be realized.

Acknowledgments

This work was sponsored by the Deputy Chief of Staff/Research, Development and Acquisition, Headquarters, U.S. Air Force, under Contract F49620-83C-0006. These results do not necessarily represent U.S. Air Force official opinion or policy.

The assistance of James Cooper, John Hanson, John Norton, Timothy Schuetze, Jack Butler, and Edward Lipp in various stages of the preparation of this paper is gratefully acknowledged.

References

[1]Easton, R.L. and Brescia, R., "Continuously Visible Satellite Constellations," NRL Rept. 6896, April 30, 1969.

[2]Walker, J. G., "Circular Orbit Patterns Providing Whole Earth Coverage," Royal Aircraft Establishment, Tech. Rept. 70211, Nov. 1970.

[3]Ballard, A. H., "Rosette Constellations of Earth Satellites," *IEEE Transactions on Aerospace and Electronic Systems,* Vol. AES16, No. 5, Sept. 1980, pp. 656, 665.

[4]Gobetz, F.W., "Satellite Networks for Global Coverage," *Advances in the Astronautical Sciences,* Vol. 9, AAS, 1963.