Traditional proof of the four-color problem.

This treatise is an attempt to traditionally prove the four-color problem.

(1)Graphed.
A map on a plane (including a sphere) consisting of n countries is made into a graph consisting of points and lines.
A country is represented by a point and a border is represented by a line connecting two points.
Treat exclave as another country.
The point of contact between countries is not a border.
If two countries are in contact with each other at multiple locations, one location will be the boundary.
The line connects two different points.
Both ends of the line cannot be the same point.
Two points at both ends of the line have different colors or color symbols.
Dots and lines do not overlap.
It doesn't overlap with itself.
There are no multiple lines between any two points.
The line is sometimes called the edge.
The line is sometimes called the side.
The line shall be a solid line.
Two points that are not connected may be connected by a dotted line.
This dotted line is called a non-connecting line.

(2)Saturation graph.
A graph in which lines that do not intersect each other on a plane are drawn as much as possible is called a saturation graph.
(fig1.)
The saturation graph divides the plane into triangles consisting of all three points and three sides.
The line of the complete graph consisting of n points is "n (n-1) / 2", and the line of the saturation graph is "3n-6".
The lines that cannot be drawn in the saturation graph (unconnected lines) are "(n-3) (n-4) / 2" by subtracting the number of lines in the saturation graph from the number of lines in the complete graph.
There are "2n-4" triangles in the saturation graph.
In this paper, we consider maps or graphs of 5 or more points in 5 countries.
The number of colors required for color coding of the saturation graph is 4 or 5.
There are four color symbols, A, B, C, and D.
If the fifth color is required, set it to E.
The countries and dots of color A are a0, a1, a2, ...
B-color countries and dots are b0, b1, b2, ...
The same applies to C color and D color.
The E color point is e0.

(3)Five-color graph generation and features.
A point with m branches is called an m branch point.
It is assumed that all saturation graphs consisting of n or less points can be painted in 4 colors.
Suppose you want to increase the points one by one and need 5 colors in a graph consisting of n + 1 points.
In that case, the first E-color point that occurs is one point e0.
The state of the graph at that time is as follows (1)-(4).

①It is a saturation graph.
In the case of an unsaturated graph consisting of n + 1 points, there is at least one polygon that is a quadrangle or larger.
If you combine the two points that face each other in the polygon, the score will be less than n + 1 and you can paint in 4 colors.
For example, in a non-saturated graph in which one line is drawn from a saturated graph, the two triangles on both sides of the drawn line become one quadrangle, and the two points facing each other in the quadrangle are combined. This makes a graph of n points, and only 4 colors are required for coloring, and even if the combination of 2 points is restored, the 4 colors remain.

②There are no 3 or 4 branch points, and all consist of 5 or more branches.
If there are 3 branch points, 4 colors will be used as an n-point saturation graph if the 3 branch points and 3 branches are removed, and 4 colors will be sufficient even if the removed points and branches are returned.
Even if the union is returned, only 4 colors are needed.

③There is one or more 5 branch points.
There are at least one or more five branch points in a saturation graph consisting of all five or more branches without three or four branch points.

④There is one E-color point.
Increase the points one by one.
Suppose that a graph that requires 5 colors appears in a saturation graph with n + 1 points.
In that case, the last added point e0 is set as the fifth color E.
If you look at the saturation graph, you can think of any point as the last added point.

(4)Point e0.
Suppose that one point e of color E appears in the saturation graph of n + 1 points.
Any point on the saturation graph can be e0.
In this paper, one of the five branch points is set to e0.
In other words, any of the 5 branch points is regarded as the n + 1th point added last.
Only the n + 1th 1 point e0 is set as the 5th color E.
e0 must be connected to at least 4 points a0, b0, c0, d0 of 4 different colors A-D with 4 branches L1-L4.
All points other than e0 are painted in 4 colors A-D.

(5)Color gamut and airspace.
Five virtual closed areas corresponding to five colors that do not overlap with the points and lines of the saturation graph are provided on the plane to correspond to the five colors.
(fig2.)
The area is called the color gamut, and the color gamut corresponding to the A color is called the A color gamut.
The same applies to the B, C, D, and E colors.
The color gamut is a virtual area that collects points of the same color.
Even if the point e0 of the fifth color E appears in the middle of increasing the number of points, only one point appears at the beginning of the appearance, so the picture showing the range of the color gamut may be omitted.
The color gamuts do not overlap.
There are no holes in the color gamut, and there is no other color gamut in the holes.
The color gamut does not come into contact with itself.
The color gamut may join with other color gamuts.
(fig3.)
The area on the plane excluding the color gamut is called the airspace.
Airspace S is seamlessly one.

(6)Point movement.
Initially, all dots and lines are in the airspace.
Move those points and collect points of the same color in each color range.
The movement of points is a one-way street from the airspace to the color gamut.
The point passes through the boundary between the airspace and the color gamut only once.
Enter directly into the color gamut corresponding to the color of the dots, and do not pass through other color gamuts.
Therefore, the airspace with a point must be in contact with the target color gamut.
In this paper, we consider only the form in which any two color gamuts touch at one point when joining the color gamuts.
Therefore, all color gamuts always touch the airspace.
Put the moved point inside the color gamut.
Do not place it on the border with other color gamuts or airspaces.
The reason is as follows.

①If you put it on the boundary line, it will be an obstacle to put the next point and the figure will be complicated.
(fig4.)

②When joining a color gamut and a color gamut, if the points are on the joined boundary line, it is not possible to know which color gamut it belongs to.
(fig5.)

(7)Entrainment of wire.
The movement of a point to the desired color gamut involves the point, its branches, and other lines that exist on the line of motion.
(fig6.)(fig7.)
A point can be avoided without colliding with another point, but it cannot (often) avoid a line connecting to another point.
The storage structure (fig8) in which the entrained lines are stored in the color gamut changes depending on the order in which the points are moved and the selection of the path.
However, the connection structure of the entire graph does not change.
The airspace borders on all color gamuts.
The order in which points are moved from the airspace to the color gamut does not prevent the remaining points from being placed in the desired color gamut.
The point that has entered the color gamut and settled remains at that position.
It does not go out into the airspace or go to other color gamuts.
In other words, there is no need to move from that position.
The point to be newly moved to the color gamut does not push out the point that is already in the color gamut, and the line is not unwound from the color gamut.
Note that the movement of a point may push the branch of another point q and push it into the color gamut.
To move a point, it is not necessary to push another point q.
A point is not moved by the movement of other points or lines.
The point that enters the desired color gamut stays there.
Only the branches are pushed and bent or stretched.
The length of the branch is not fixed.

(8) Separation graph A graph in which points are placed in each color gamut is called a division graph.
The following is observed in the fractional graph.

①There can be points in the decentralized graph that appear to have passed through other color gamuts.
This is because the branch at that point gets caught up in the movement of another point.
(fig9.)

②Looking at the decentralized graph, it can be regarded as generating points and lines within each color gamut and extending branches from the generated points to points in different color gamuts, instead of moving the points.
Even if the graph is generated in this way, the order in which points and lines are generated is not restricted by a specific rule.

(9)Overlay the map and graph.
Overlay a graph that represents the map with dots and lines on a map where the two countries meet at the boundary.
Observing it, the dots are within each land and do not need to be outside the land.
The line connecting the two points does not have to leave the land of the two countries.
If we consider the land as the color gamut of each color, we can consider that the two color gamuts are joined.
Considering that the graph is attached to the map, the alignment of the graph can be changed by transforming the land area of the map.
For example, a specific curved line can be made into a straight line.
The connection structure of the graph does not change even if it is transformed.
(fig10.)

(10)Kempe chain.
A partial map or subgraph in which two different colors are assigned to a land or dots are alternately called a Kempe chain.
If any two countries or two points of two different colors are selected, the Kempe chain of the two colors is either present or absent between the two countries or two points.
There can be multiple routes if they exist.
If there are multiple routes, select one route R.
The land of the countries that make up R can be regarded as part of the color gamut of each color, and the entire land can be re-divided to create a color gamut in which the two color gamuts are joined.
(fig11-fig14.)
Then, all the elements such as points and lines on the route R can be put in the joined color range.
There is no need to put it out of the joined color range.
Even if the color gamuts of the two colors are joined, they are in contact with the airspace.
Therefore, there is no problem in putting the points remaining in the airspace into each color range.
The exchange of the two colors used in the Kempe chain is called color exchange.
Color exchange extends not only to the selected route R, but also to the extent where the two colors are connected.

(11)Occurrence of point e0 of the fifth color E.
It is assumed that the point of the fifth color E is generated by the increase of the points.
In the graph at that time, the point of color E is only e0.
e0 must be connected to the other four colored dots.
There must be a subgraph like (fig15, fig16) as the subgraph structure of the graph.

(12)Invert the inside and outside of the graph.
Here, the inside and outside of the graph are inverted around e0.
The figure is called an inverted figure.
Let e0 be the wall of color E, and the branches of e0 come out of the wall.
(fig17.)
Four color gamuts corresponding to A-D colors are provided in the wall.
In order for the point e0 to be the color E, the points a0, b0, c0, and d0 at the tips of the branches L1-L4 from e0 must be in four different color gamuts.
The points in each of the four color gamuts are connected by a line to the points in the other color gamuts.
Do not consider the connection between points in the adjacent color gamut when viewed from e0.
Adjacent color gamut points are of course different colors and are ignored as they may or may not be connected by lines.
What should be considered is the connection between points in opposite color gamuts.
In the inverted view, the roots of L1-L4 are oriented north, south, east, and west, and the wall is divided into four sections: northeast, southeast, northwest, and southwest.

(13)Presence or absence of Kempe chain.
In the inverted view, the case where the Kempe chain does not exist (or does not form) and the case where it exists (or does) are considered separately between a0-b0 or c0-d0 that straddle two opposite color gamuts.

①When there is no Kempe chain in both a0-b0 and c0-d0 (fig18) a0, b0, c0 and d0 are directly connected to e0 at L1-L4.
Since e0 has 5 branch points, there is L5.
If one end ai of L5 is in the A color gamut, the other end must be directly connected to the opposite northeast or southeast wall e0 on the B color gamut side.
Therefore, the points in the A color gamut and the B color gamut are left as they are.
Perform color exchange of C-D color with either Kempe chain of C-D color connected to c0 or d0, and make c0 and d0 the same color.
As a result, the colors of the points directly connected to the wall e0 become the three colors A, B, and C (or D), and the wall e0 can be changed to the C or D color instead of the color E.

The same applies when one end of L5 is in any of the B, C, and D color gamuts.

②When there is a Kempe chain at a0-b0 or c0-d0.
The kempe chains of a0-b0 and c0-d0 cannot exist at the same time.
Here, it is assumed that a0-b0 has a Kempe chain.
When there is a Kempe chain between two points, it is not limited to only one route due to the branching and merging of lines.
In that case, select one route R.
The land of four countries corresponding to a0, b0, c0, and d0, which are directly connected to e0, is regarded as four color gamuts.
Next, the A color gamut and the B color gamut of the root R are joined, and the area of the joined color gamut is re-divided.
Then, all the points and lines on R can be put in the joined color gamut.
There are two ways to use L5.

②-1 When L5 is parallel to R.
(fig9).
If one end of L5 is the ai point in the A color gamut, L5 must be directly connected to the northeast or southeast wall.
The A color gamut is divided into two upper and lower parts by R and r.
Since R and r cannot cross, whether the direct connection destination is northeast or southeast depends on whether ai is above or below.
c0-d0 has no Kempe chain.
If you exchange colors with a CD color Kempe chain containing c0 (or d0), the points directly connected to e0 will be 3 colors (A, B, D (or C) colors), so e0 will be C color (or D color). Can be.
The same applies when L5 comes out of the B color gamut.

②-2 When L5 tries to cross R.
(fig20).
If one end of L5 is a ci point in the C color gamut, L5 needs to be directly connected to the southeast or southwest wall, but it is blocked by R and r of a0-b0 and cannot be reached.
If you exchange colors with the CD color Kempe chain including d0 and change d0 to C color, the points directly connected to e0 will be 3 colors (A, B, C colors), so e0 can be changed to D color. ..
The same applies when L5 comes out of the D color gamut.

(14)Straight line rigidity and color gamut minimization.
However, the following can be considered here.

(1) When the C color gamut extends and crosses the line r of b0-e0, which is an extension of R, and one end ci of L5 is included in the tip of the tongue (fig21).

(2) When the line of b0-r0 extends and divides the C color gamut, and one end ci of L5 is included in the tip of the tongue (fig22).
Both are shapes in which dots are generated in the color gamut beyond r.
This is because I think of it as a graph consisting of points and lines.
Implicitly, (1) the shape and size of the color gamut are free, and (2) the line can be freely extended and bent.
(If a0 and b0 are in direct contact with e0 and there is no r (fig23), the color gamut and color gamut may join but do not overlap, so color gamut C and color gamut D are It is completely blocked by the junction of color gamut A and color gamut B.
Also, R does not go out of the junction of color gamut A and color gamut B.
As a result, the problems (1) and (2) above do not occur.
The above problems (1) and (2) can be solved by setting the following two conditions.
The first condition.
Minimize the color gamut.
Second condition.
Let L1-L4 be a straight rigid body.
As the first condition, the color gamut is minimized.
The color gamuts do not overlap, and they are not large enough to overlap the lines going to other color gamuts in L1-L4.
Even if it is minimized, if there is an area, any number of points can be inserted, so there is no need to increase it.

As the second condition, L1-L4 is made into a linear rigid body.
One ends of L1-L4 gather at e0, and since L1-L4 do not overlap, they form a cross-like radial shape while keeping their angles.
L1-L4 does not bend and pass through other color gamuts.
This cross-shaped subgraph consisting of L1-L4 and e0, a0, b0, c0, and d0 is called a nucleus.
(fig24) (fig16).
Each of e0's L1-L4 goes to the desired color gamut, and there is no need to go to the other color gamut.
However, L5 must go to one of the color gamuts separated by one, not to the adjacent color gamut.
To do that, you have to bend.
It is assumed that only L1-L4 is a straight rigid body, and all other lines (including L5) are flexible, bend and extend.
Even so, there is no need to change the connection structure of the graph.
Even under this condition, the line from cj must reach either the left or right wall (southeast or southwest) at the base of d0.
However, it is not possible due to the presence of R and r, and only the northeast and northwest walls are reachable.
L5 can only be parallel to R and r.
If a graph that requires 5 colors exists or can be created, it should be possible to draw a graph that requires 5 colors even under these 2 conditions.
However, that is not possible as mentioned above.
The core that requires 5 colors does not occur.
Due to the above, graphs or maps that require 5 colors cannot be drawn on a flat surface.

(14)Deformation of the land Note that the linear rigidity of L1-L4 can be explained as follows.
Overlay the subgraph consisting of e0, a0, b0, c0, d0, L1-L4 on the corresponding submap.
Assuming that the partial map and the partial graph are sticky, the sticky graph can be deformed and the curve can be straightened by deforming the partial map.
Such transformation does not change the connection structure between countries.
It is the four branches L1-L4 directly connected to e0 that are made into a straight rigid body.
The other lines (L5 connected to e0 and other lines) are said to bend freely.
e0, a0, b0, c0, d0, L1-L4 were collectively called the nucleus.
Seen from the whole graph, the nucleus is a local entity.
Even if it occurs at the n + 1 point, there is only one, not multiple, so there is no new problem because L1-L4 does not bend.
The core is very simple and intuitive.
It is clear that adopting the idea of a nucleus does not affect the connection structure of the graph.

FIG.1

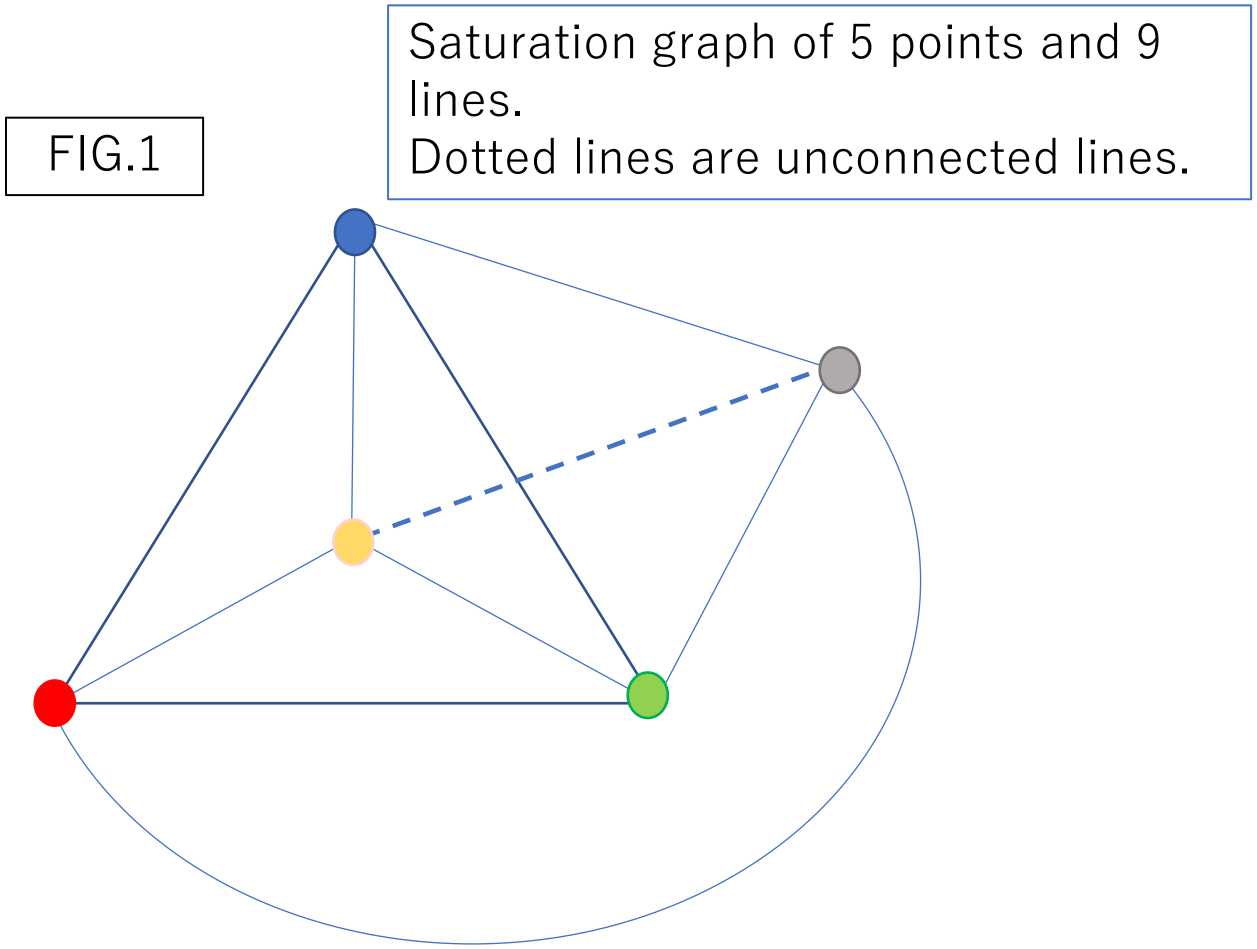

Saturation graph of 5 points and 9 lines.
Dotted lines are unconnected lines.

FIG.2

Create 5 color gamuts.
E is the color gamut of the fifth color.
Airspace S outside the color range.

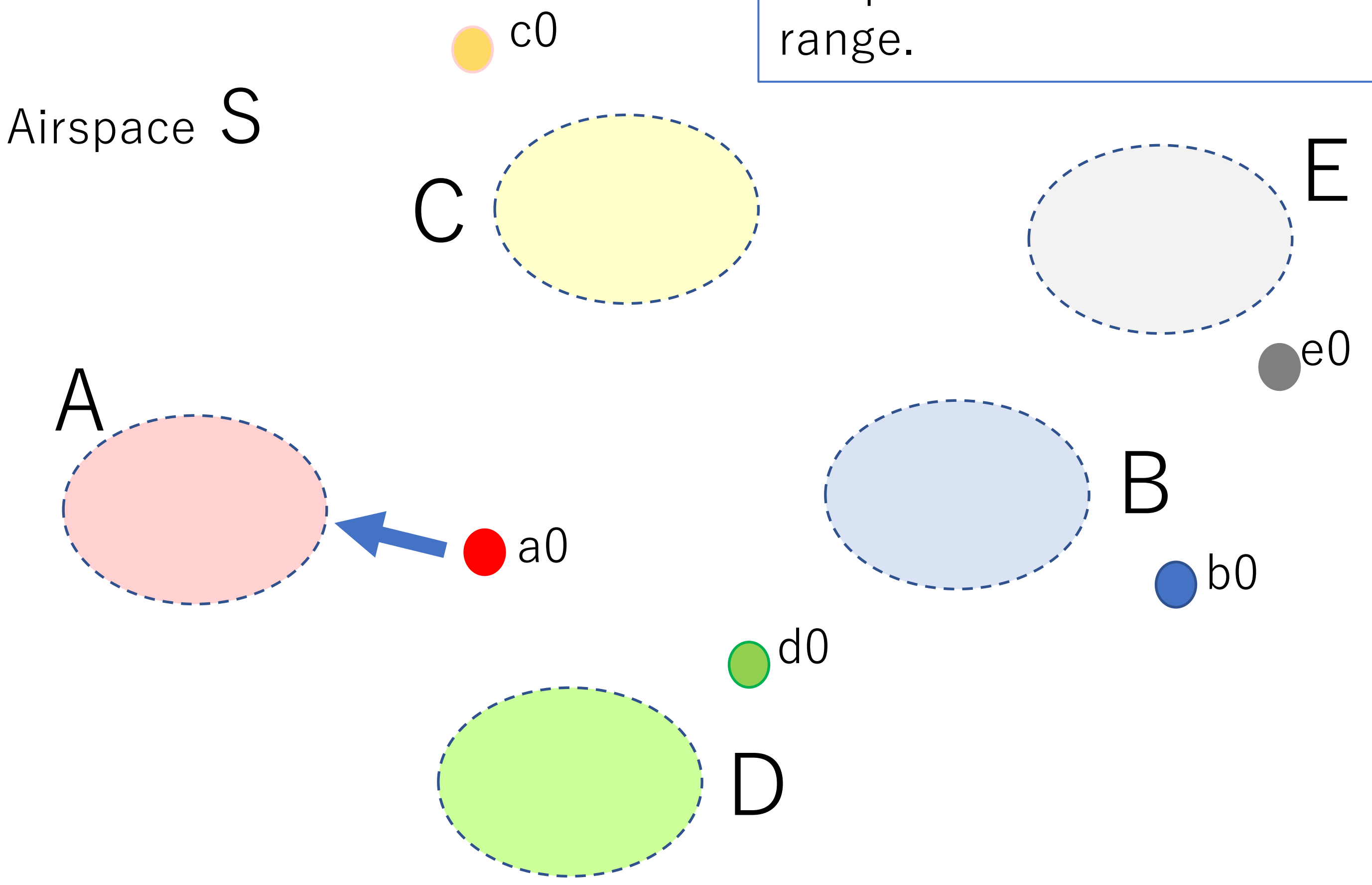
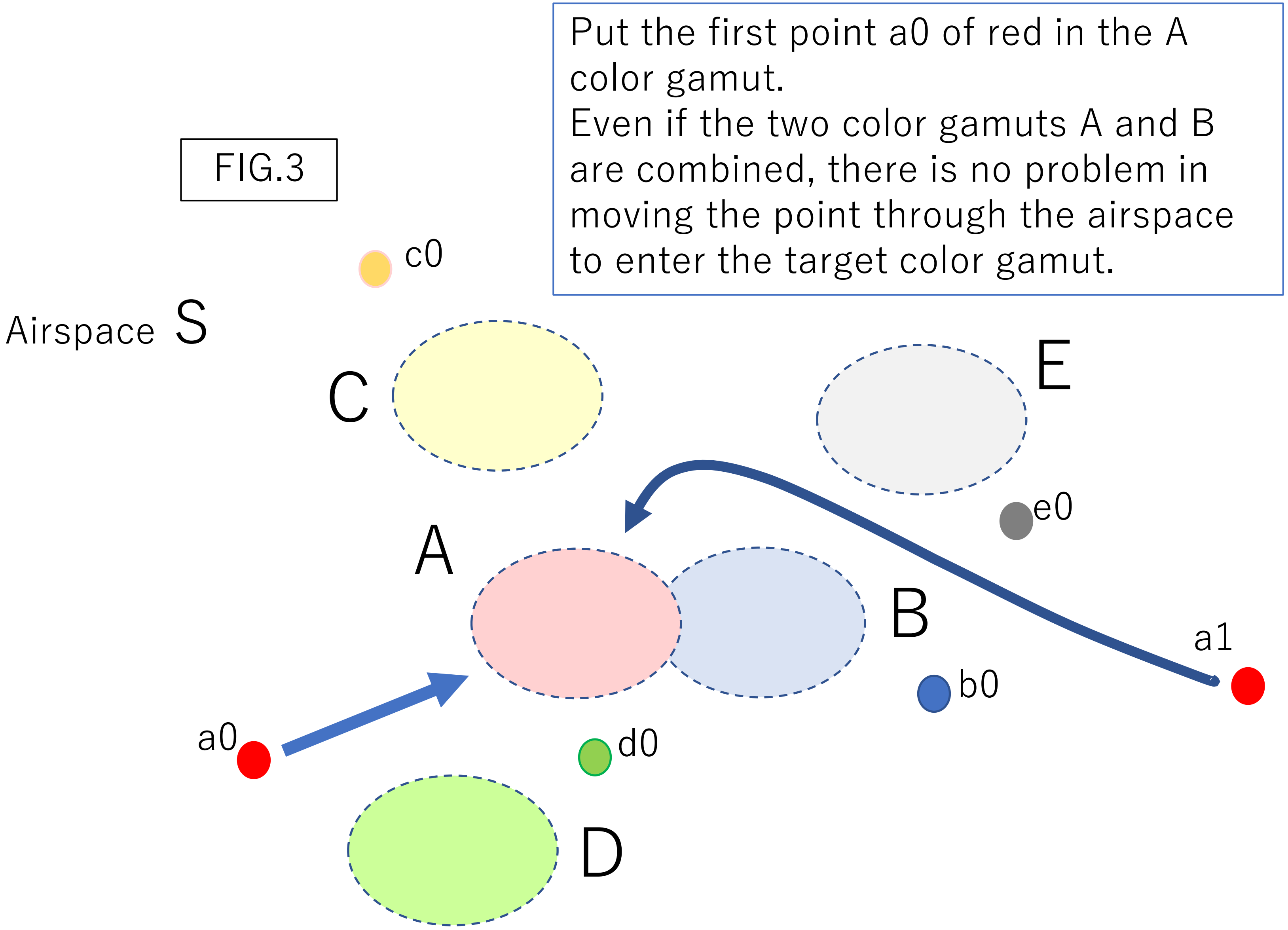
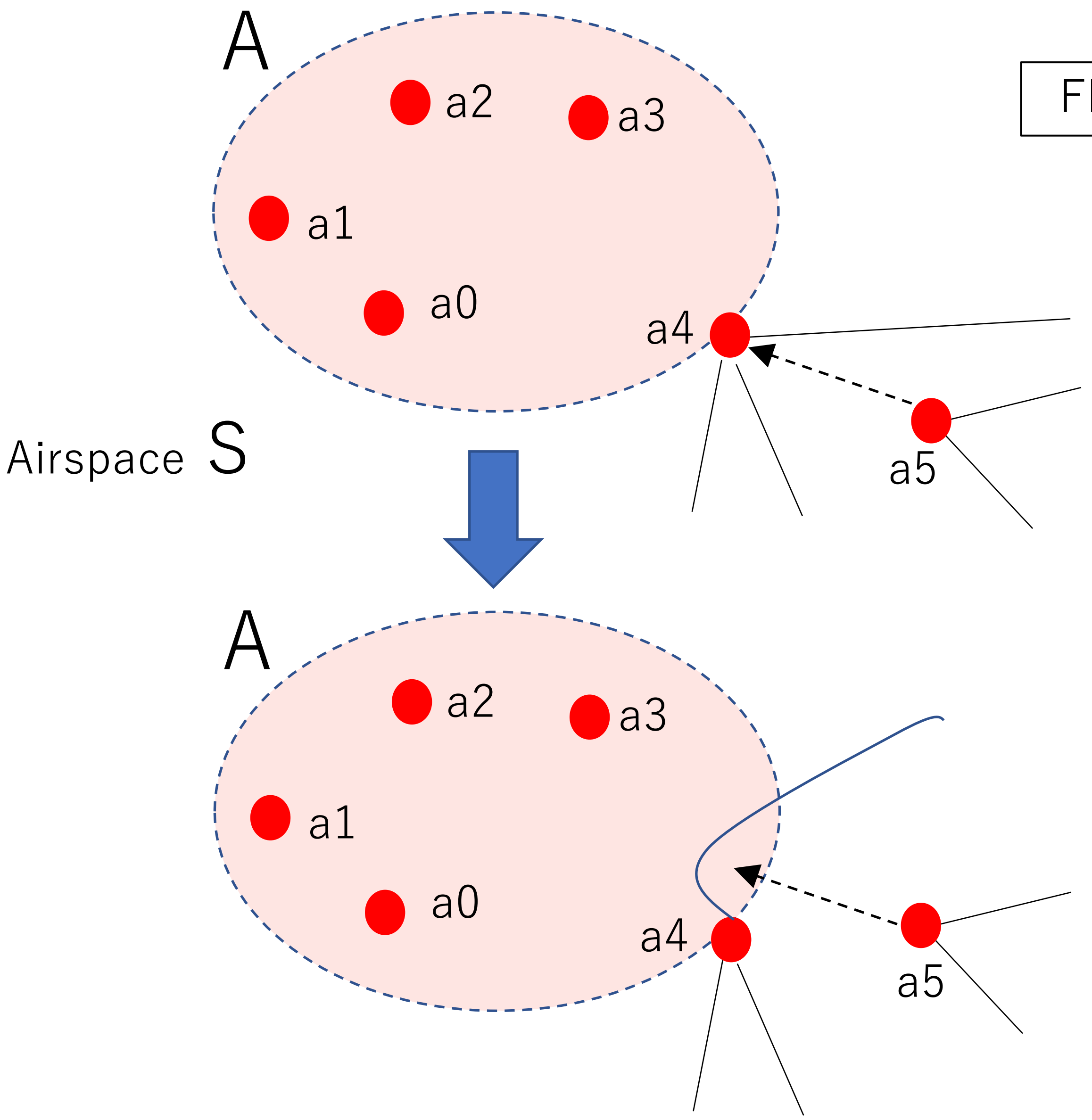


FIG.3



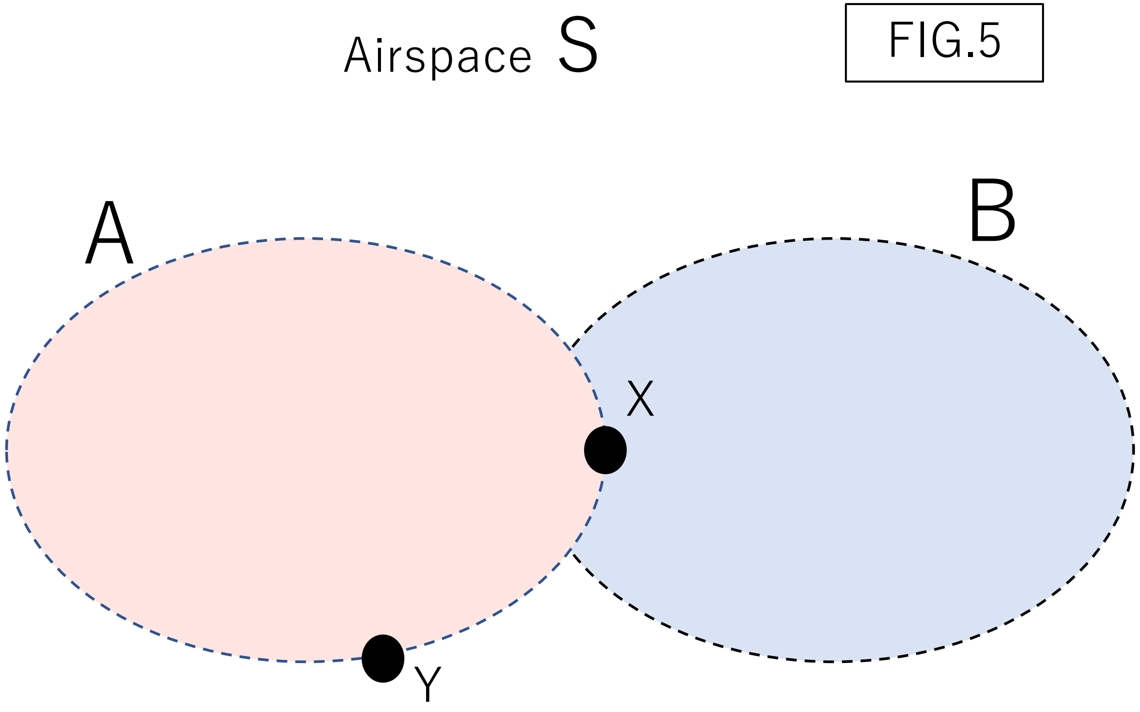
Put the first point a0 of red in the A color gamut.
Even if the two color gamuts A and B are combined, there is no problem in moving the point through the airspace to enter the target color gamut.
c0
Airspace S
C
E
e0
A
B
a1
b0
a0
d0
D

FIG.4



If you place it on the border of the color gamut like point a4, a5 will not fit smoothly, so you have to bend the line of a4 as shown in the figure.
The figure becomes complicated.

FIG.5



It is unknown which area the point X on the boundary between the A color gamut and the B color gamut belongs to.
The point Y at the boundary between the airspace S and the A color gamut may be considered to belong to the A color gamut, but it is clearer to put it in the A color gamut.

FIG.6

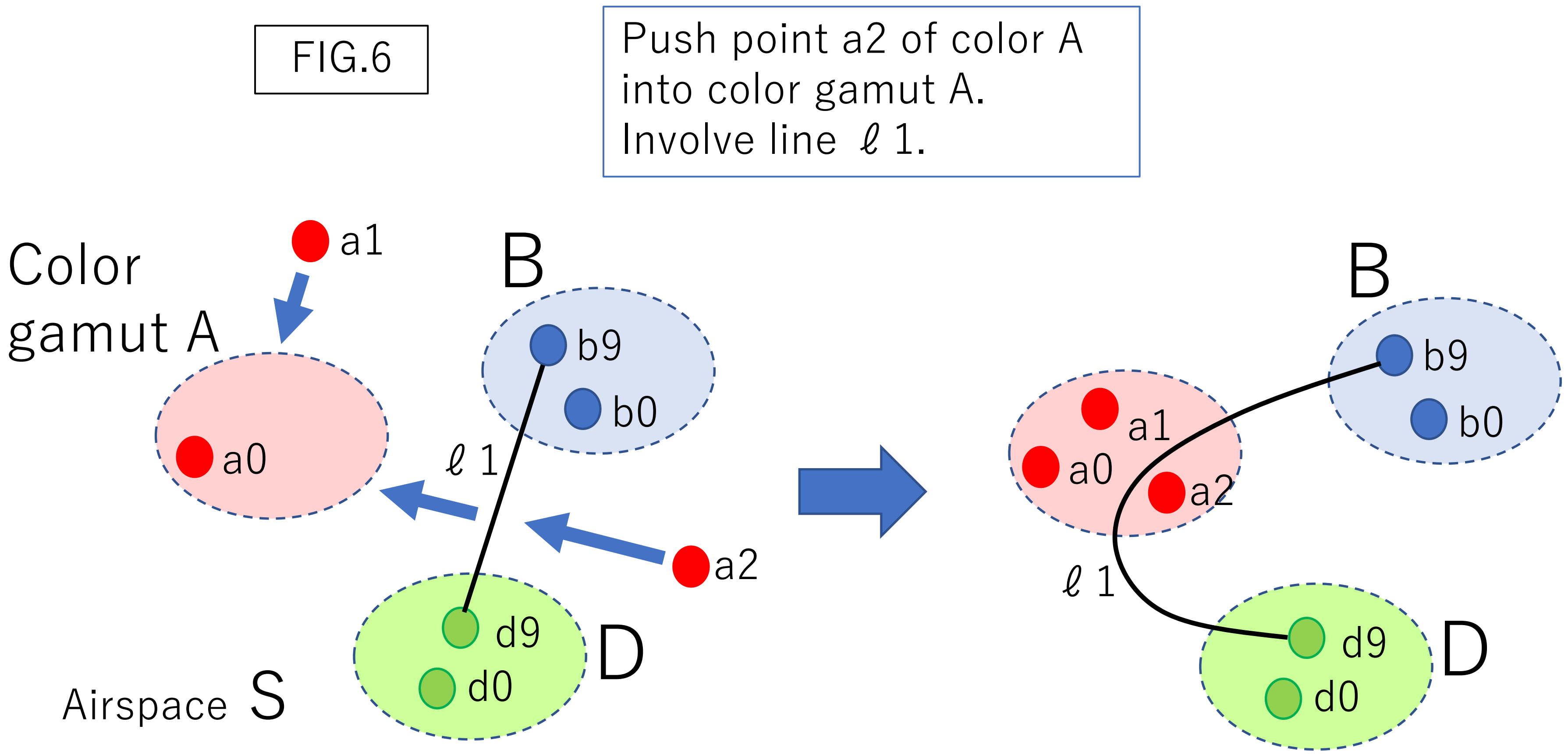

Push point a2 of color A
into color gamut A.
Involve line ℓ1.
a1
Color
gamut A
B
b9
b0
a0
ℓ1
a2
d9
d0
D
Airspace S
B
b9
b0
a1
a0
a2
ℓ1
d9
d0
D

FIG.7

Collect points of the same color in the color gamut (Example 1). E is the color gamut of the fifth color.

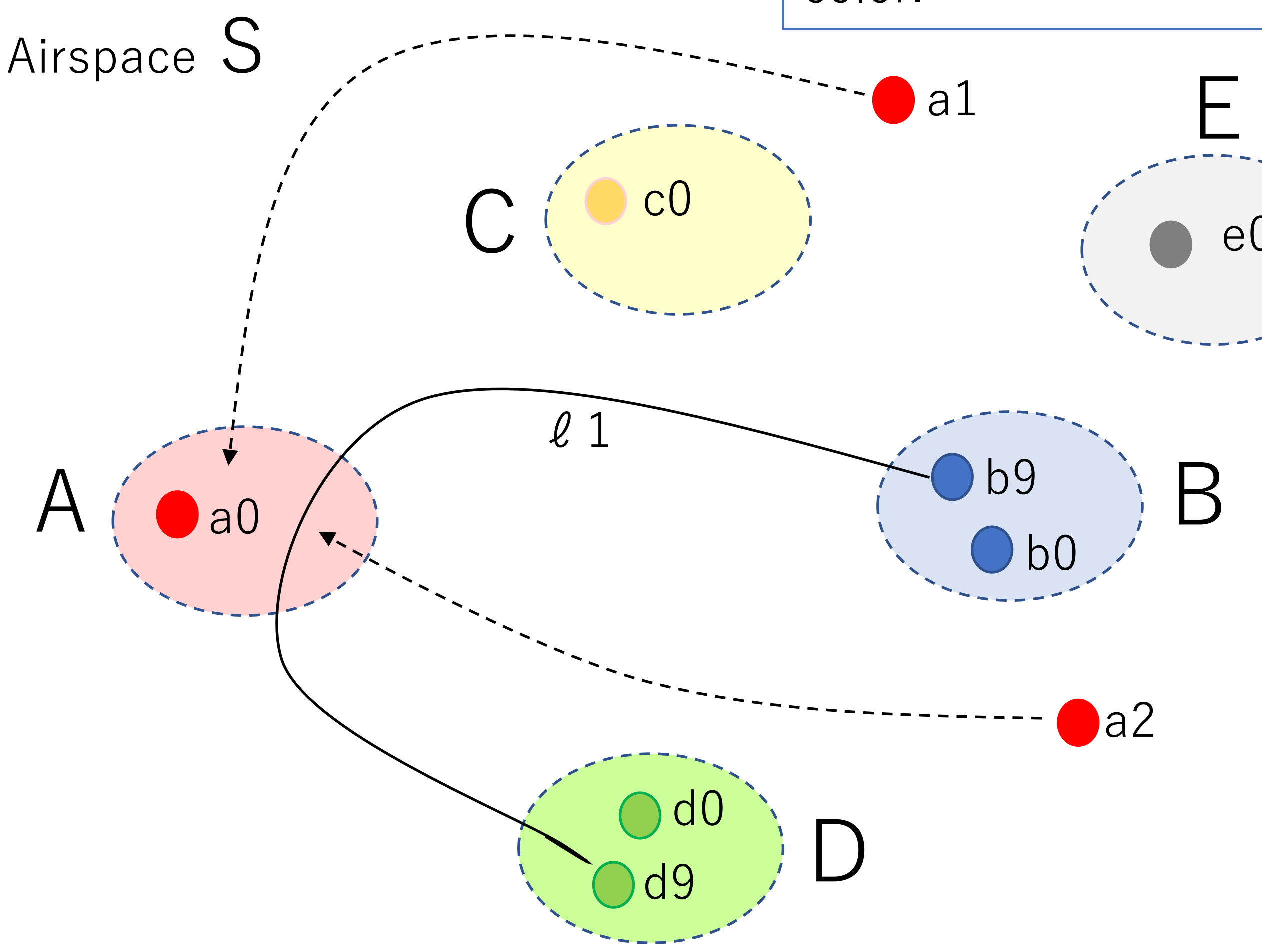
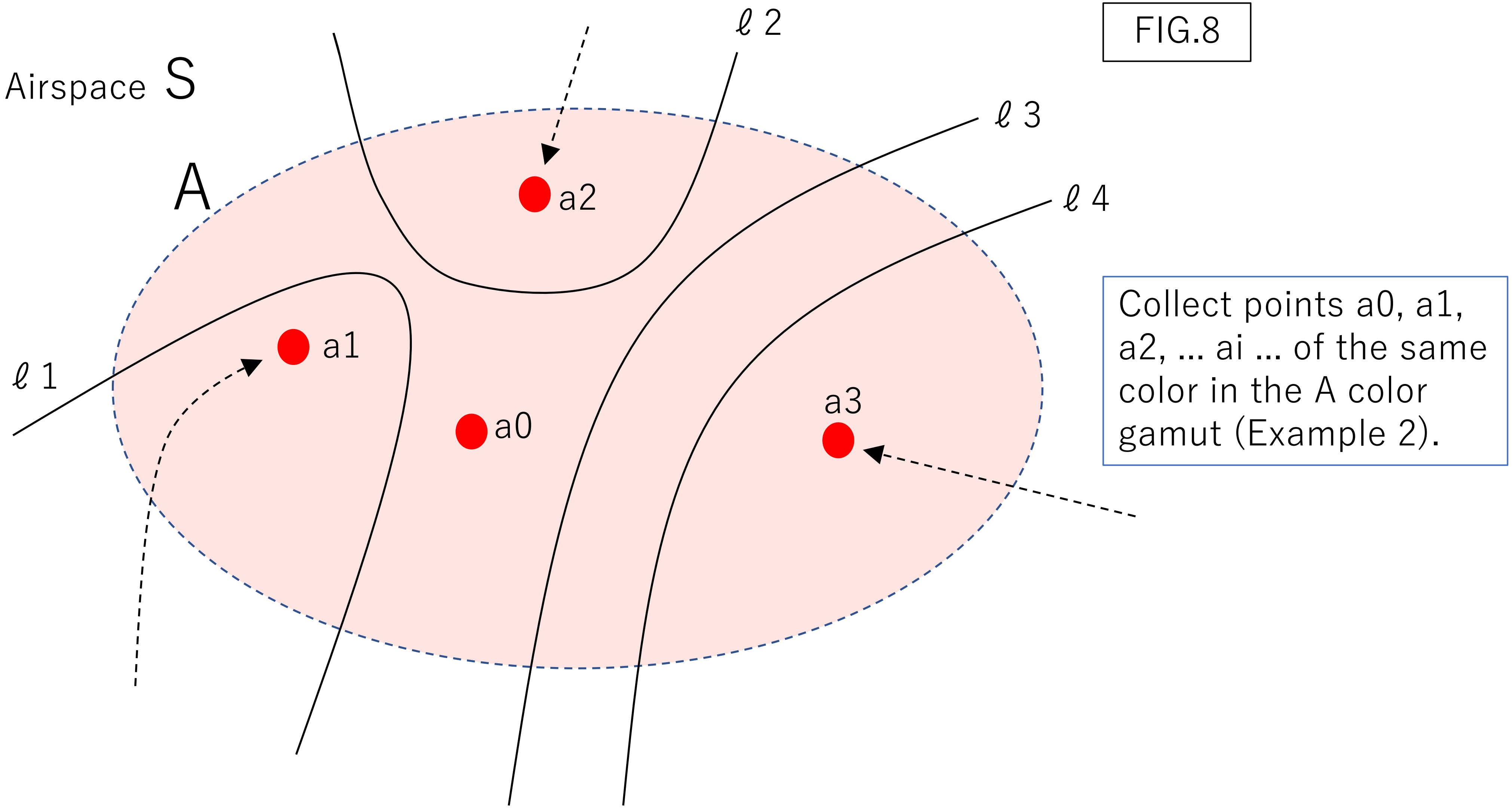


FIG.8


Airspace S
A
ℓ 1
ℓ 2
ℓ 3
ℓ 4
a0
a1
a2
a3
Collect points a0, a1, a2, ... ai ... of the same color in the A color gamut (Example 2).

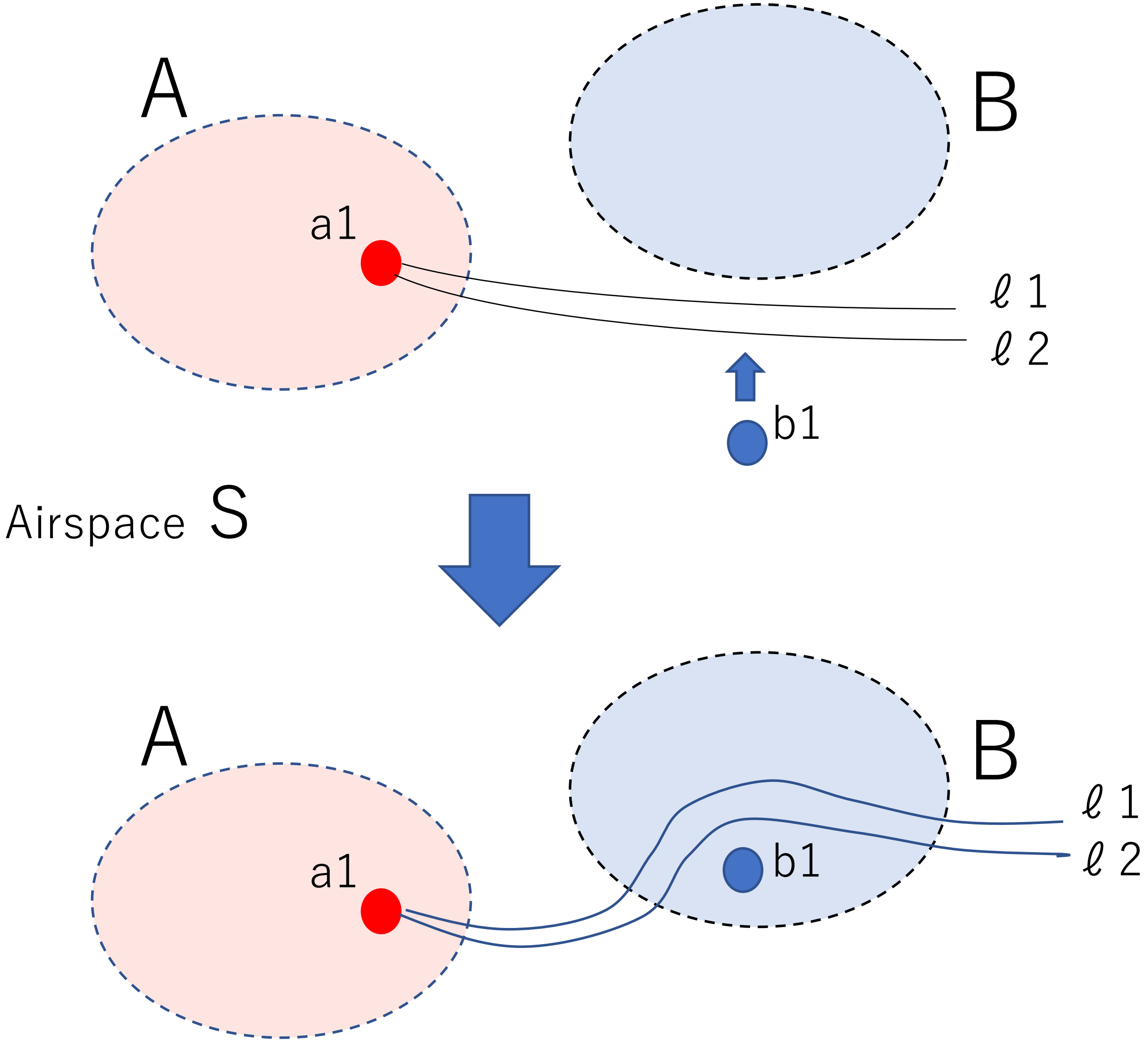


FIG.9

The lines ℓ 1 and ℓ 2 connected to the point a1 are pushed into the B color gamut by moving the point b1 to be put into the B color gamut.
This makes it appear that the point a1 and the lines ℓ 1 and ℓ 2 have passed through the B color gamut.

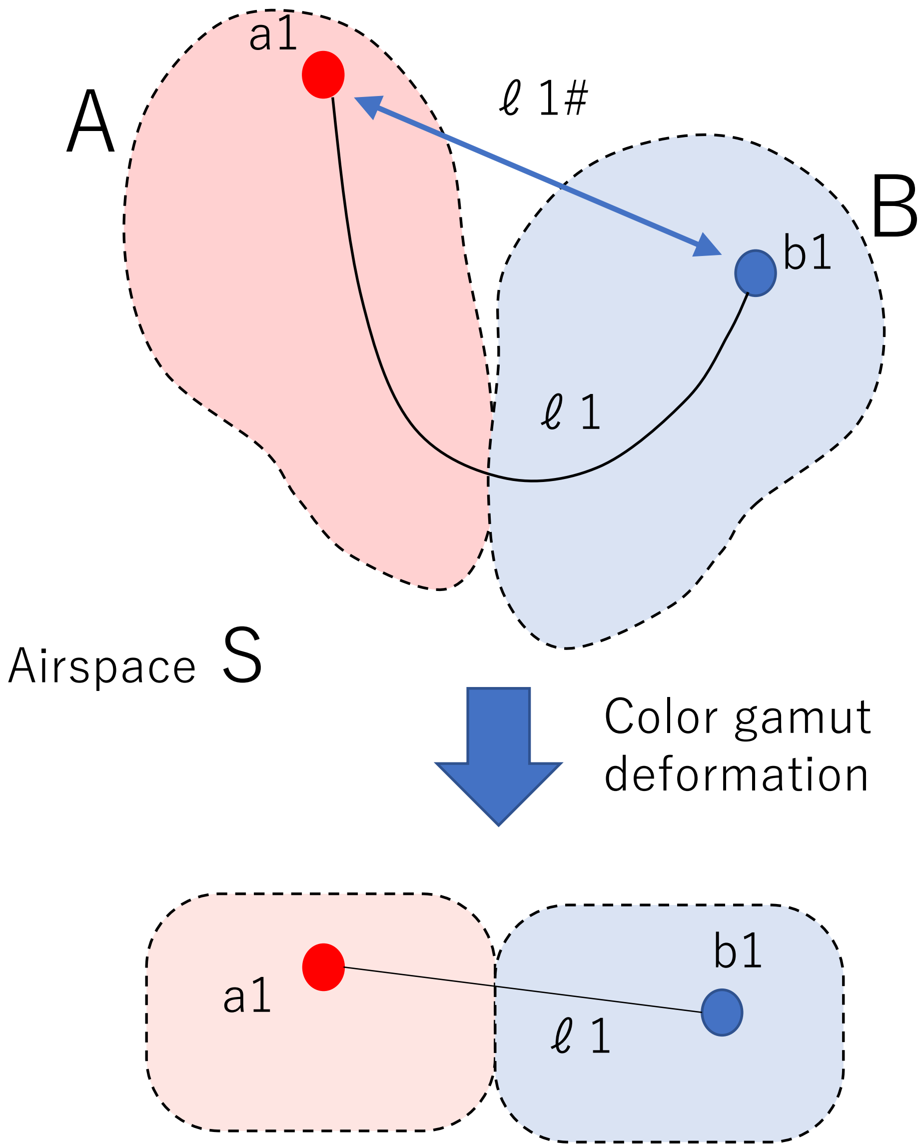
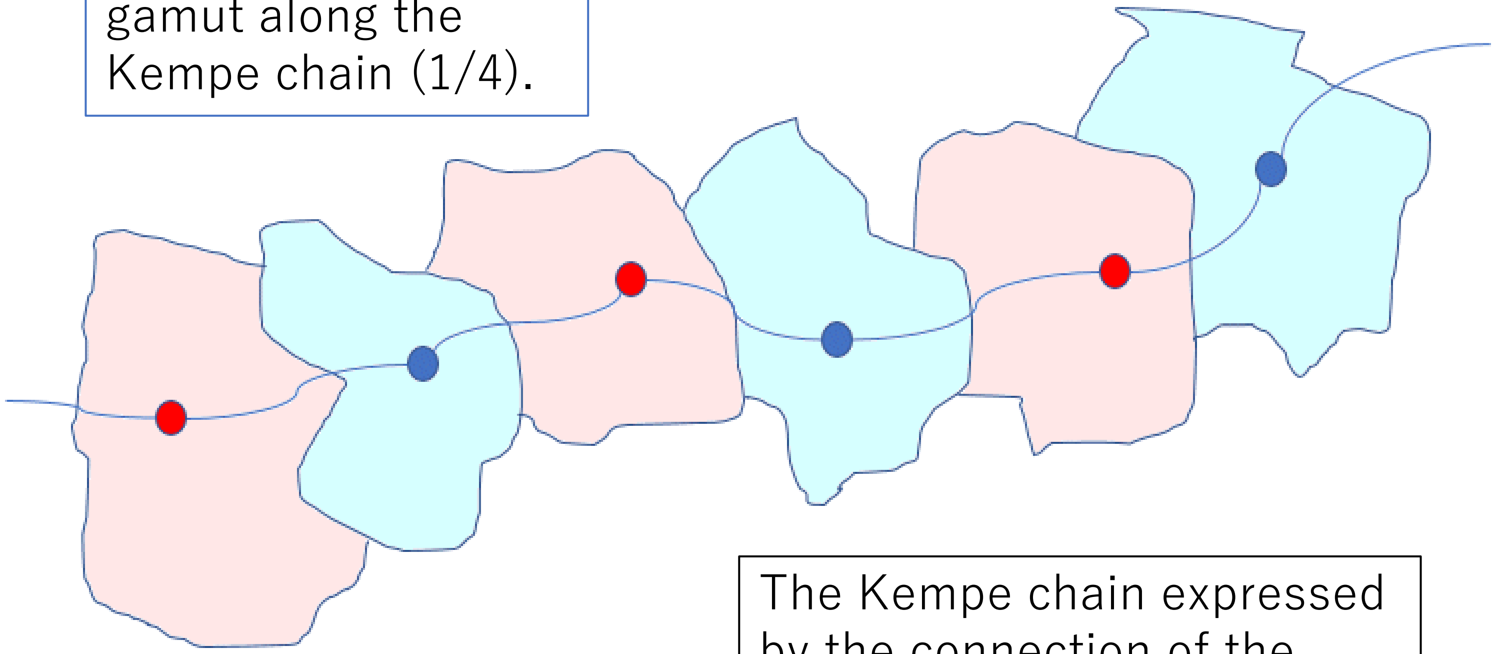


FIG.10

The line ℓ 1 connecting the points a1 and b1 bends and passes through the boundary between the A color gamut and the B color gamut. Line ℓ 1 does not have to go out into airspace S.
However, in this figure, the line ℓ 1 #, which is a straight line of the line ℓ 1, appears in the airspace S.
The color gamut can be transformed. ℓ 1 can be linearized by appropriately transforming the A color gamut and the B color gamut.

An example of creating a color gamut along the Kempe chain (1/4).

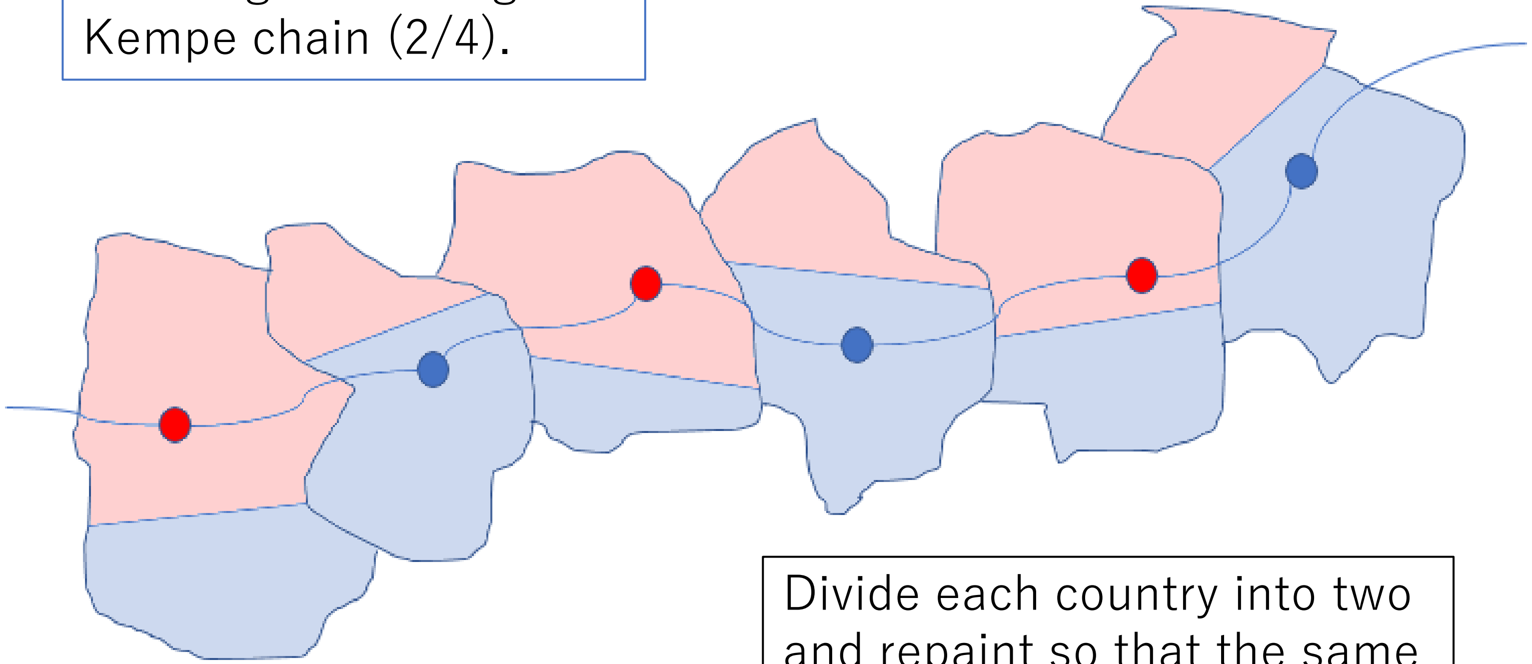
FIG.11



The Kempe chain expressed by the connection of the land and the Kempe chain expressed by the graph are overlapped and drawn.

FIG.12

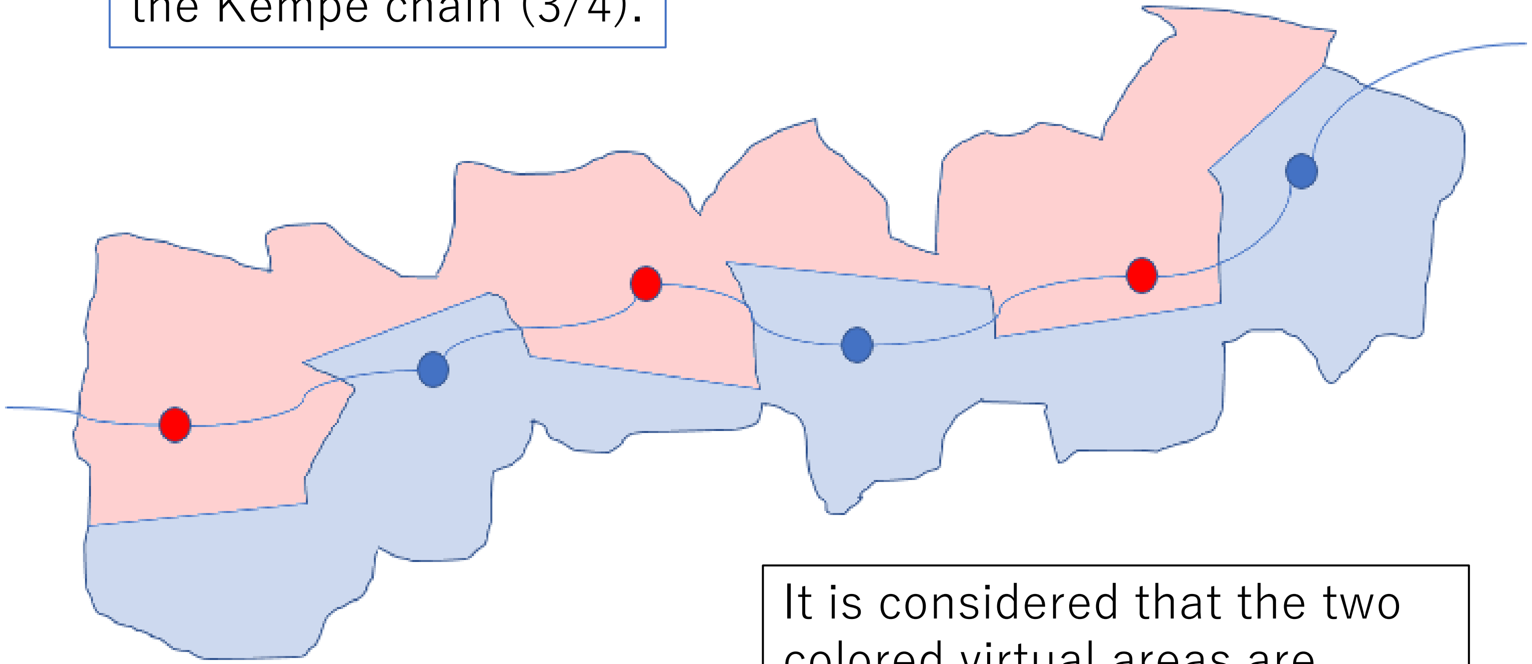
An example of creating a color gamut along the Kempe chain (2/4).



Divide each country into two and repaint so that the same colors are connected.

An example of creating a color gamut along the Kempe chain (3/4).

FIG.13

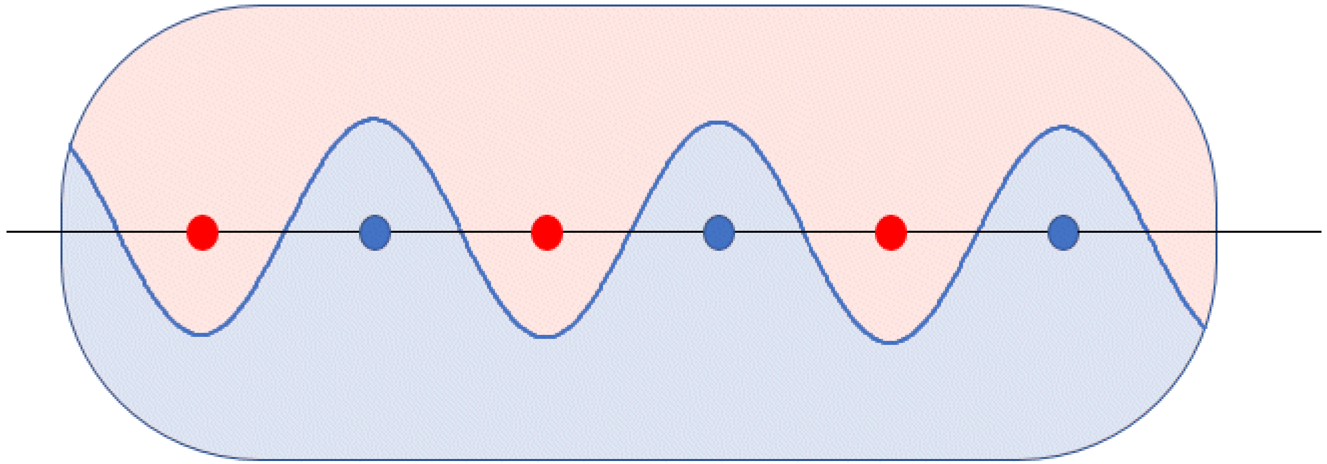


It is considered that the two colored virtual areas are joined by connecting the land of the same color.

An example of creating a color gamut along the Kempe chain (4/4).

FIG.14

Two examples of redrawing FIG.13



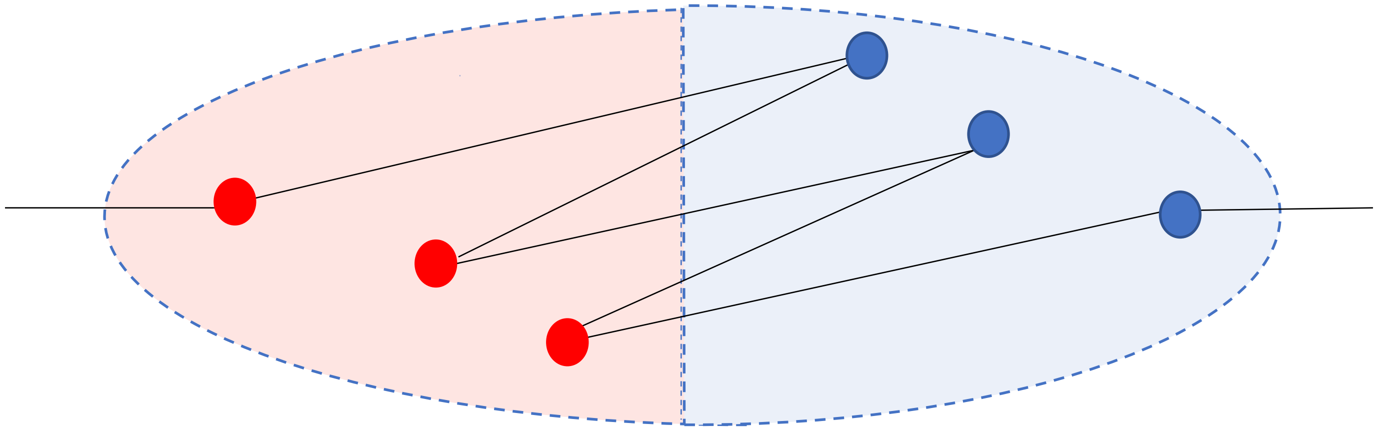
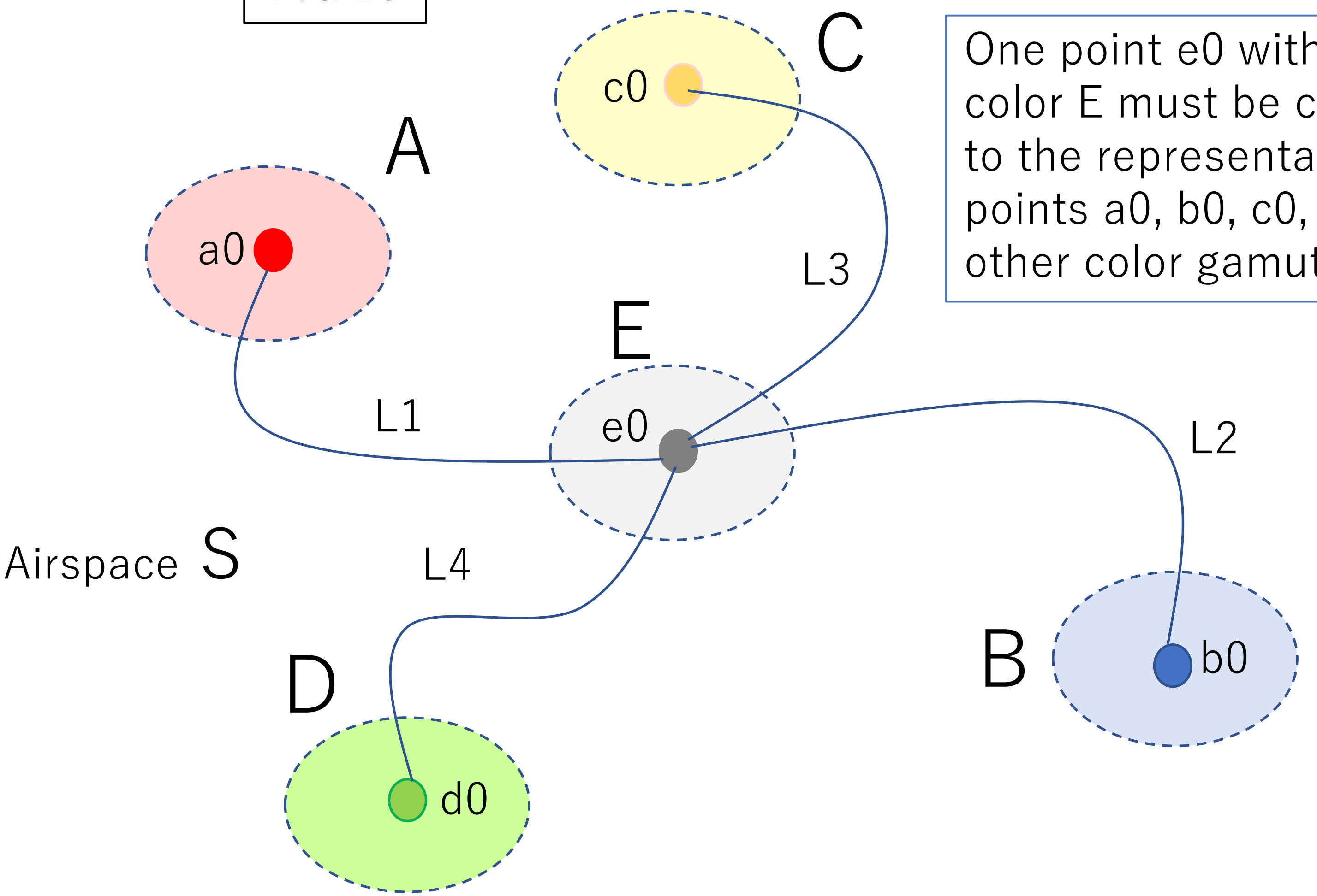
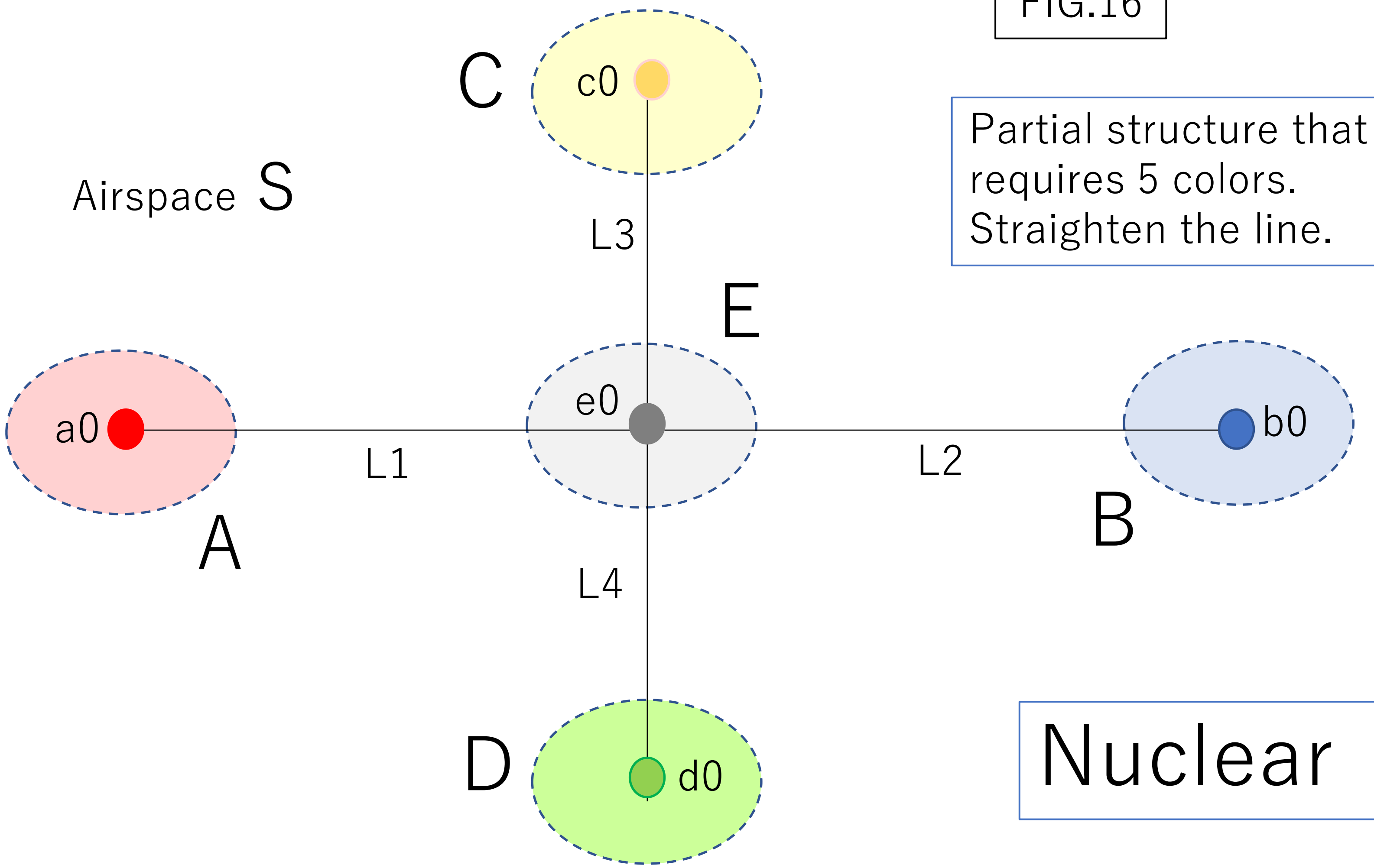


FIG.15



One point e0 with the fifth color E must be connected to the representative points a0, b0, c0, d0 of other color gamuts.

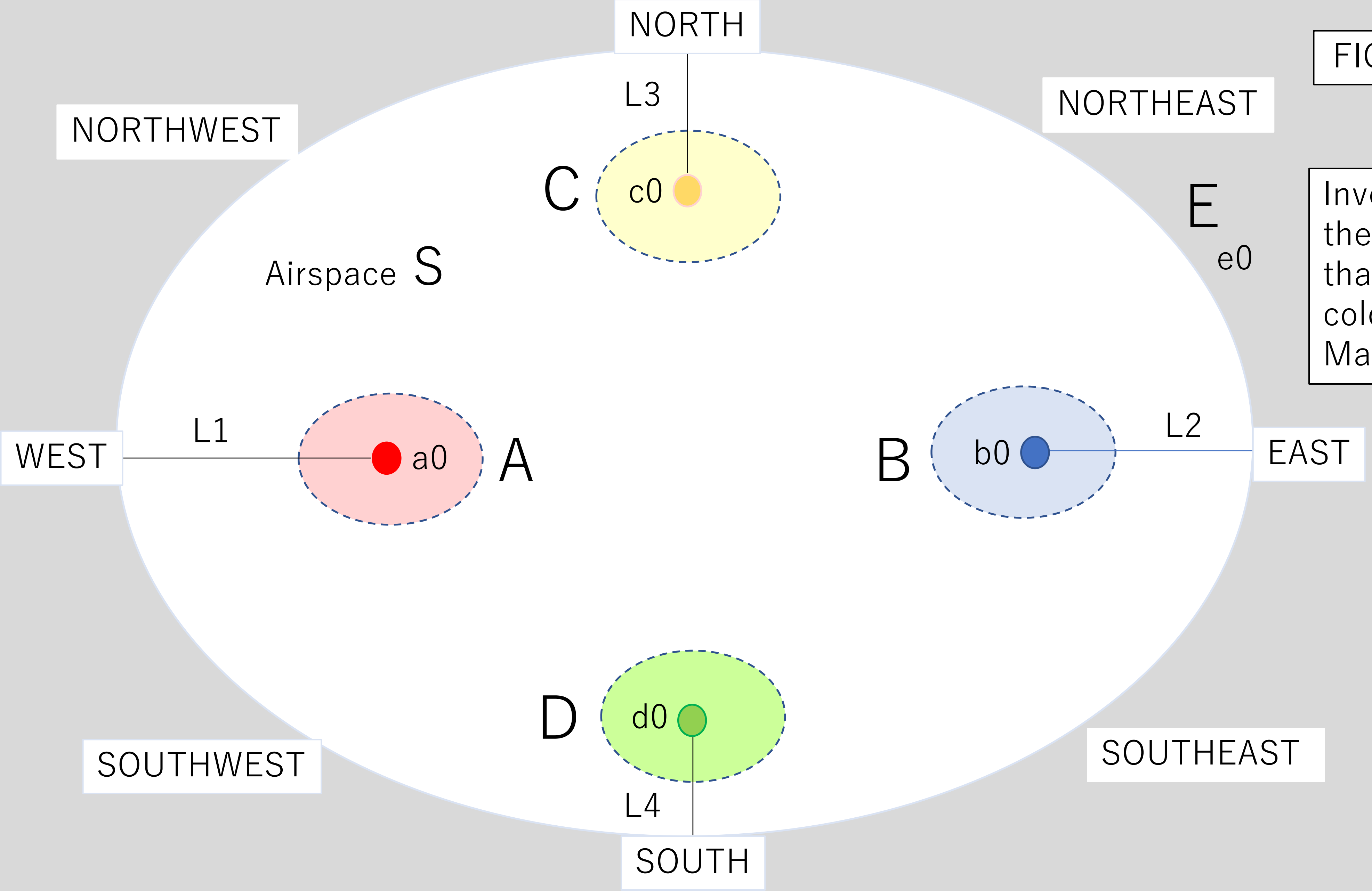
FIG.16



Partial structure that requires 5 colors. Straighten the line.

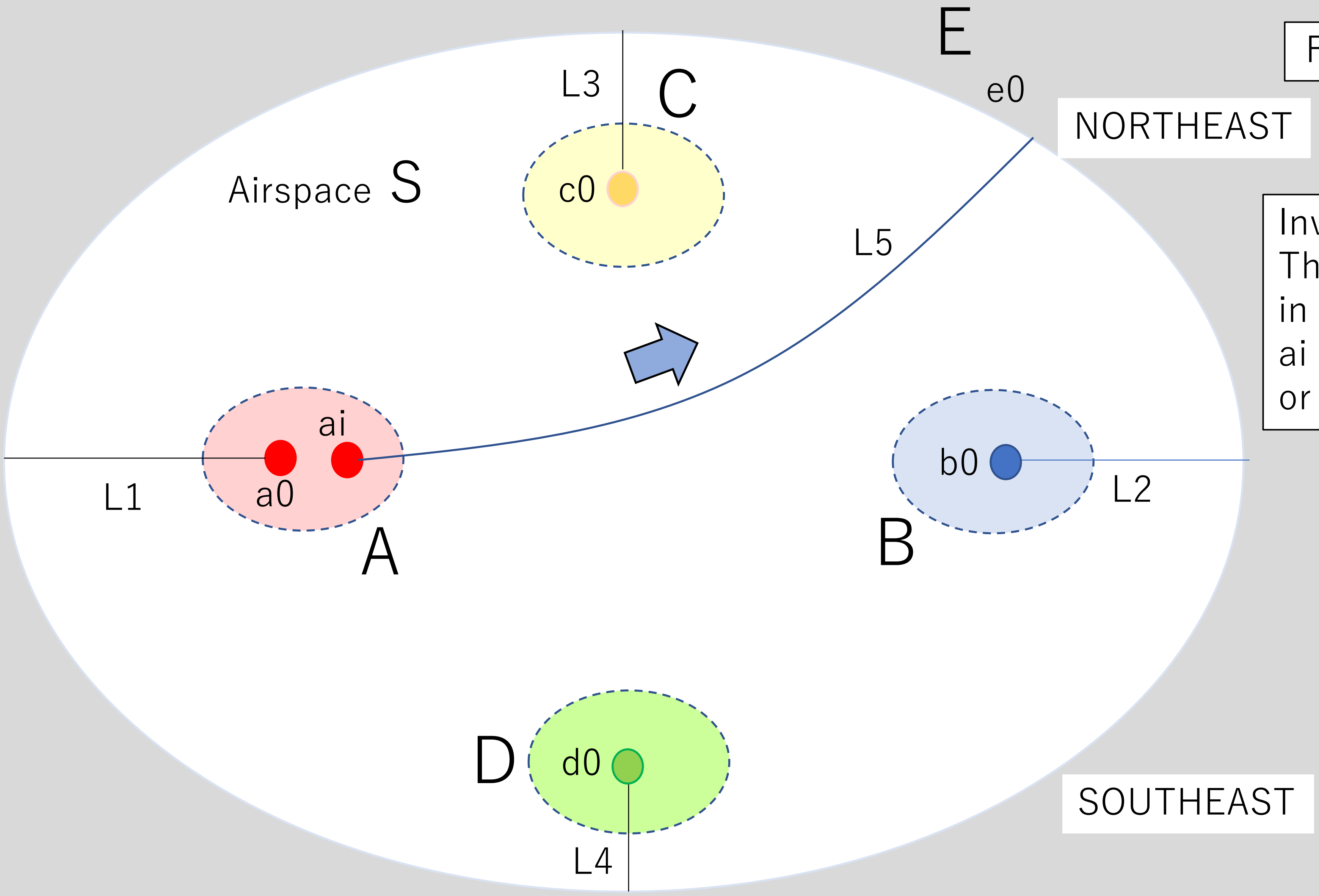
Nuclear

FIG.17



Inversion graph of the partial structure that requires 5 colors.
Make a direction.

FIG.18



Inverted graph.
There is no kempe chain in a0 to b0 and c0 to d0.
ai can reach the northeast or southeast wall of e0.

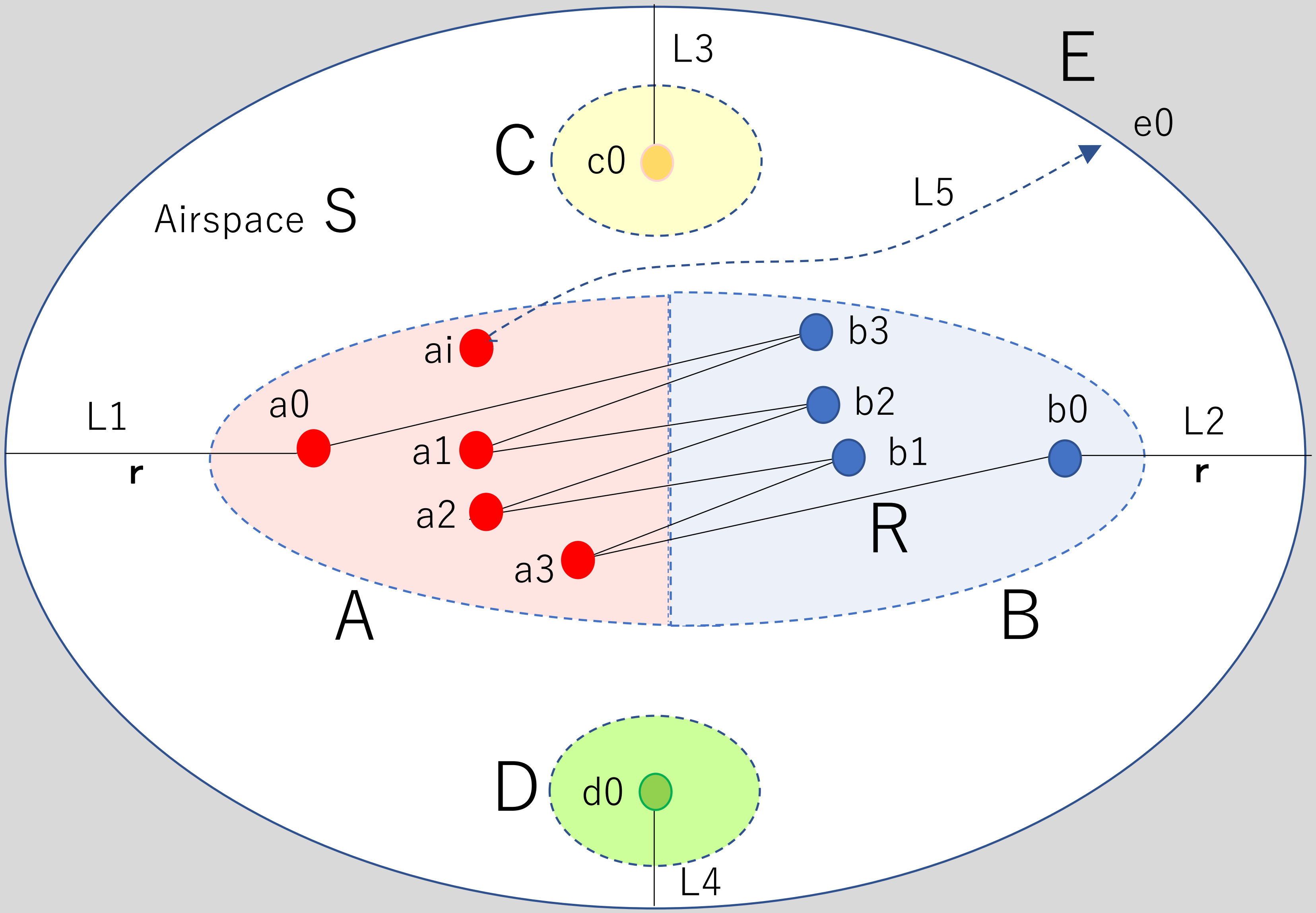


FIG.19

The color gamuts A and B are joined, and there is a Kempe chain R in a0 to b0.
Both ends a0 and b0 of R are connected to e0 by a line r.
Line L5 of ai cannot be said to be inaccessible to the northeastern wall, but not to the southeastern or southwestern wall.

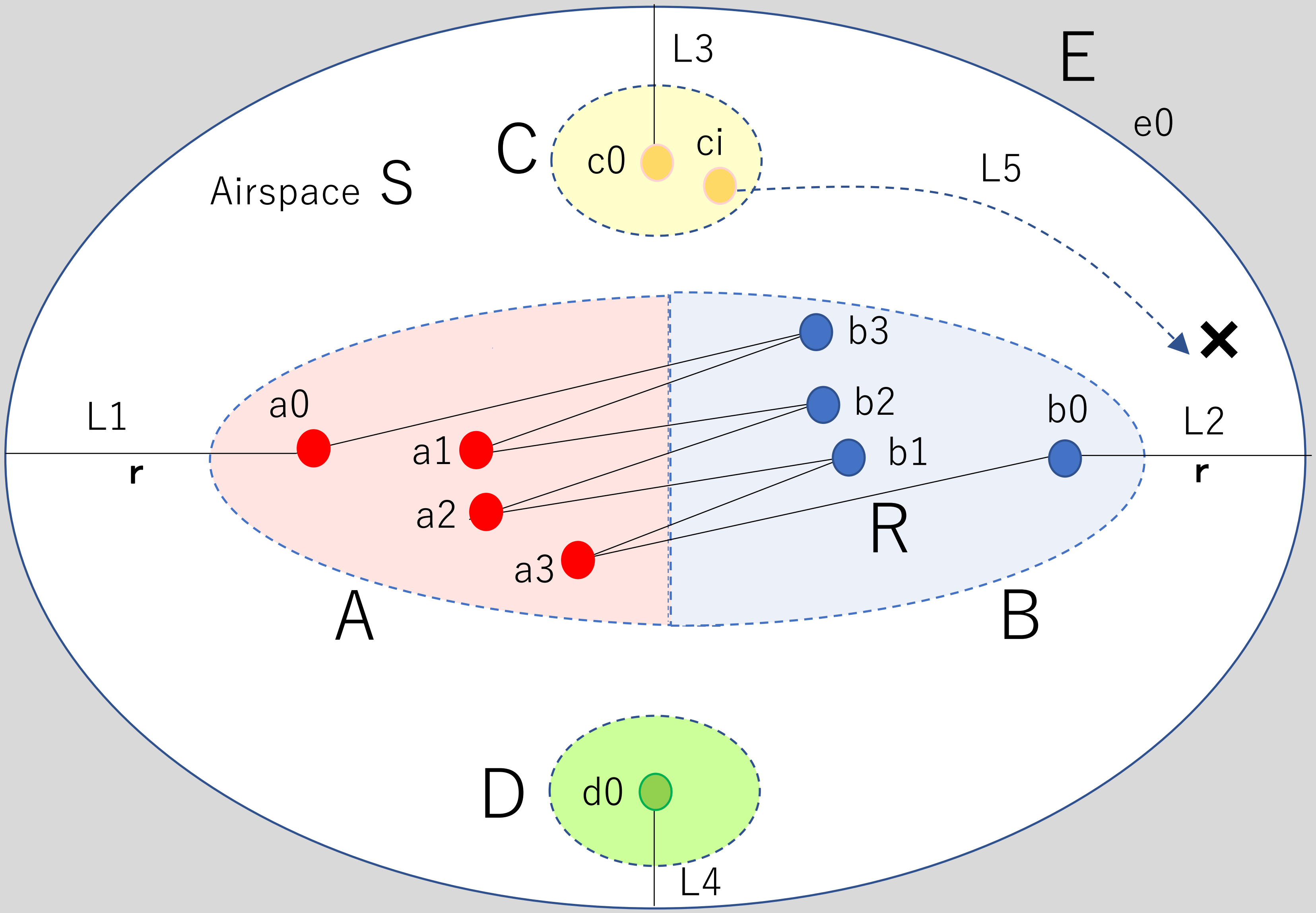


FIG.20

The color gamuts A and B are joined, and there is a Kempe chain R in a0 to b0. Both ends a0 and b0 of R are connected to e0 by a line r.
The ci line L5 is inaccessible to the southeast and southwest walls and points in the D color gamut.

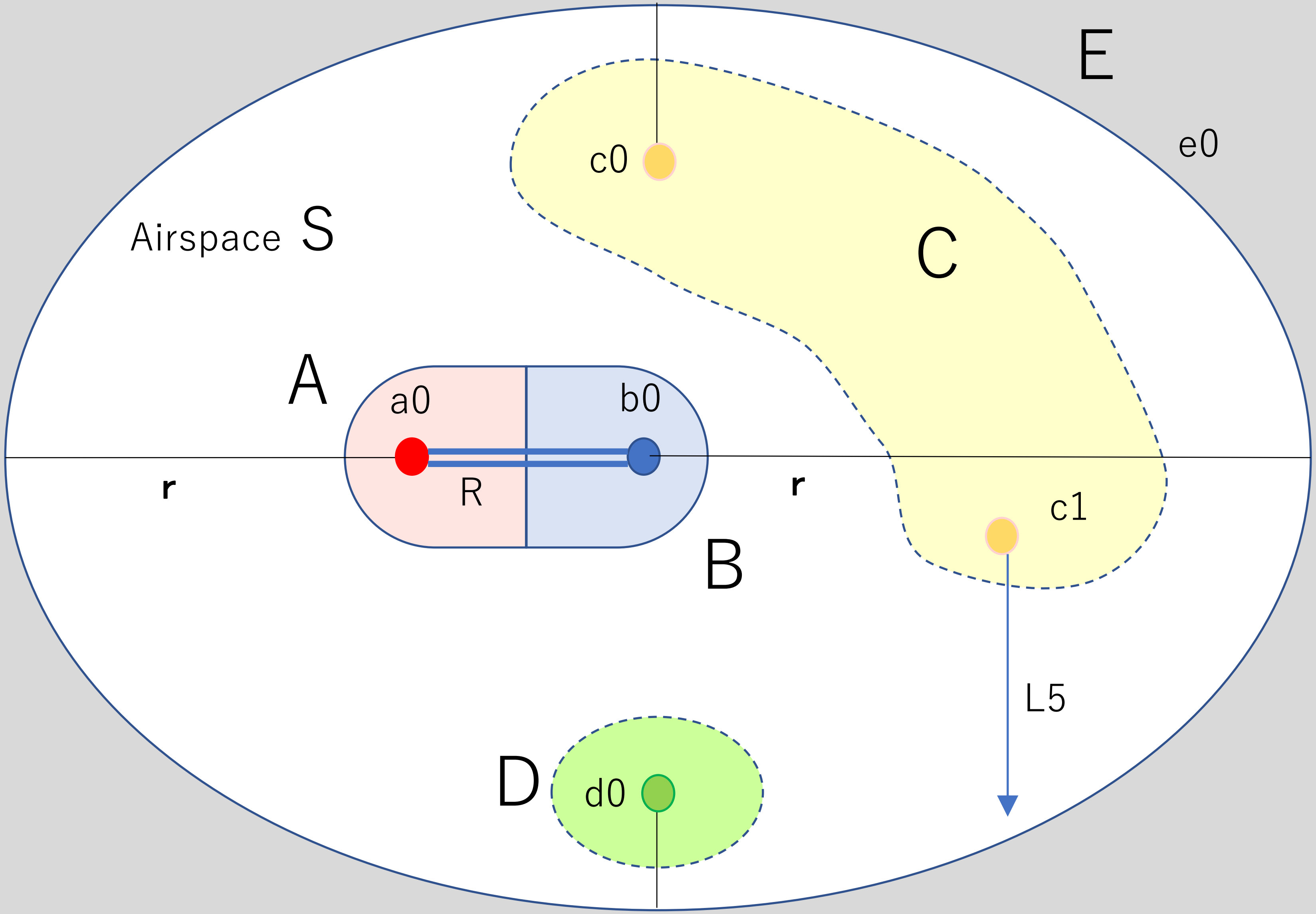


FIG.21

The color gamuts A and B are joined, and there is a Kempe chain R (drawn with a double line) in a0 to b0.
Both ends a0 and b0 of R are connected to e0 by a line r, but since the color gamut C extends and c1 is below r, c1 may be connected to the wall on the color gamut D side by L5.

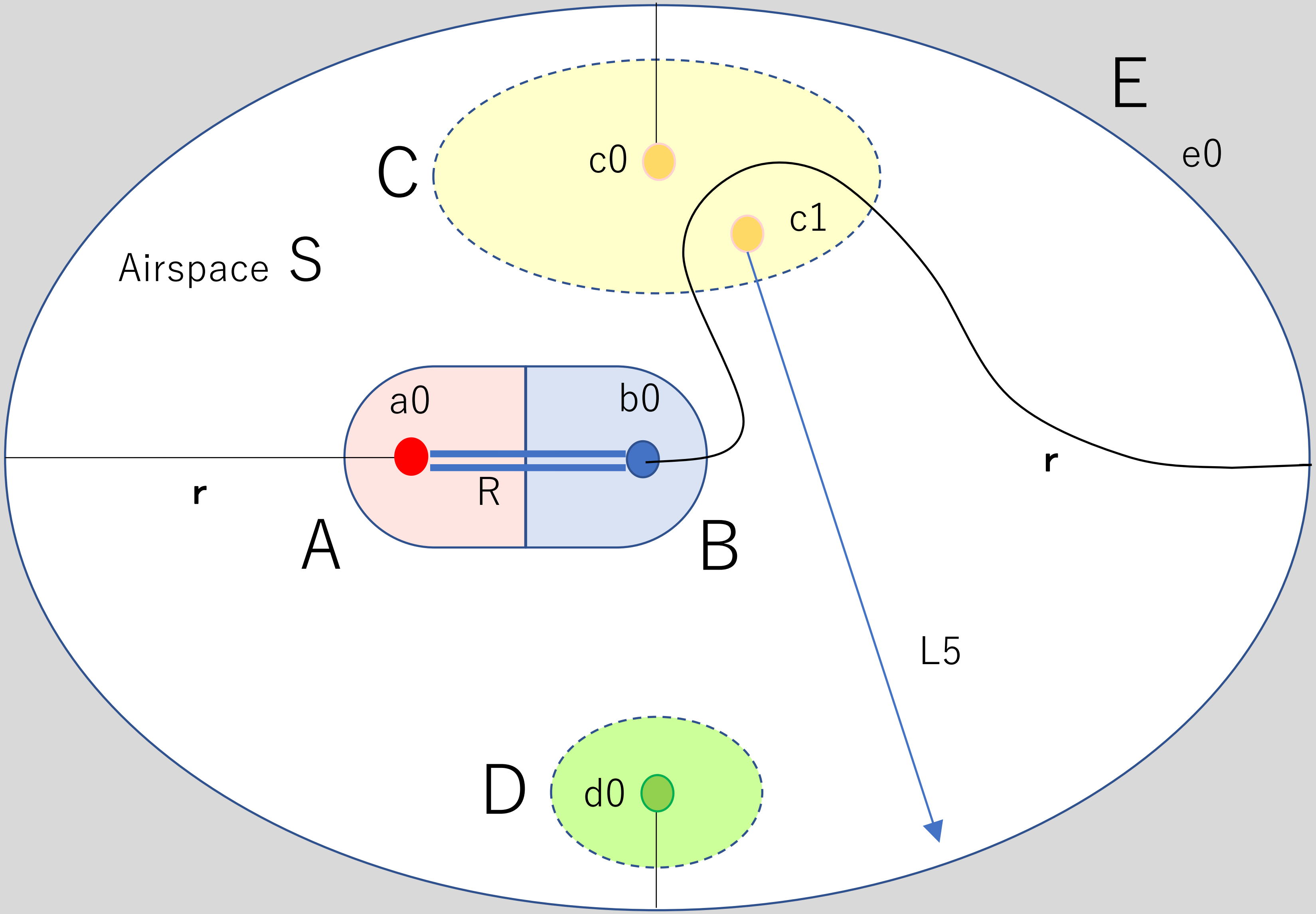


FIG.22

The color gamuts A and B are joined, and there is a Kempe chain R (drawn with a double line) in a0 to b0.
Both ends a0 and b0 of R are connected to e0 by a line r.
But since r passes through the color gamut C and c1 is below r, it may be connected to the wall on the color gamut D side by L5.

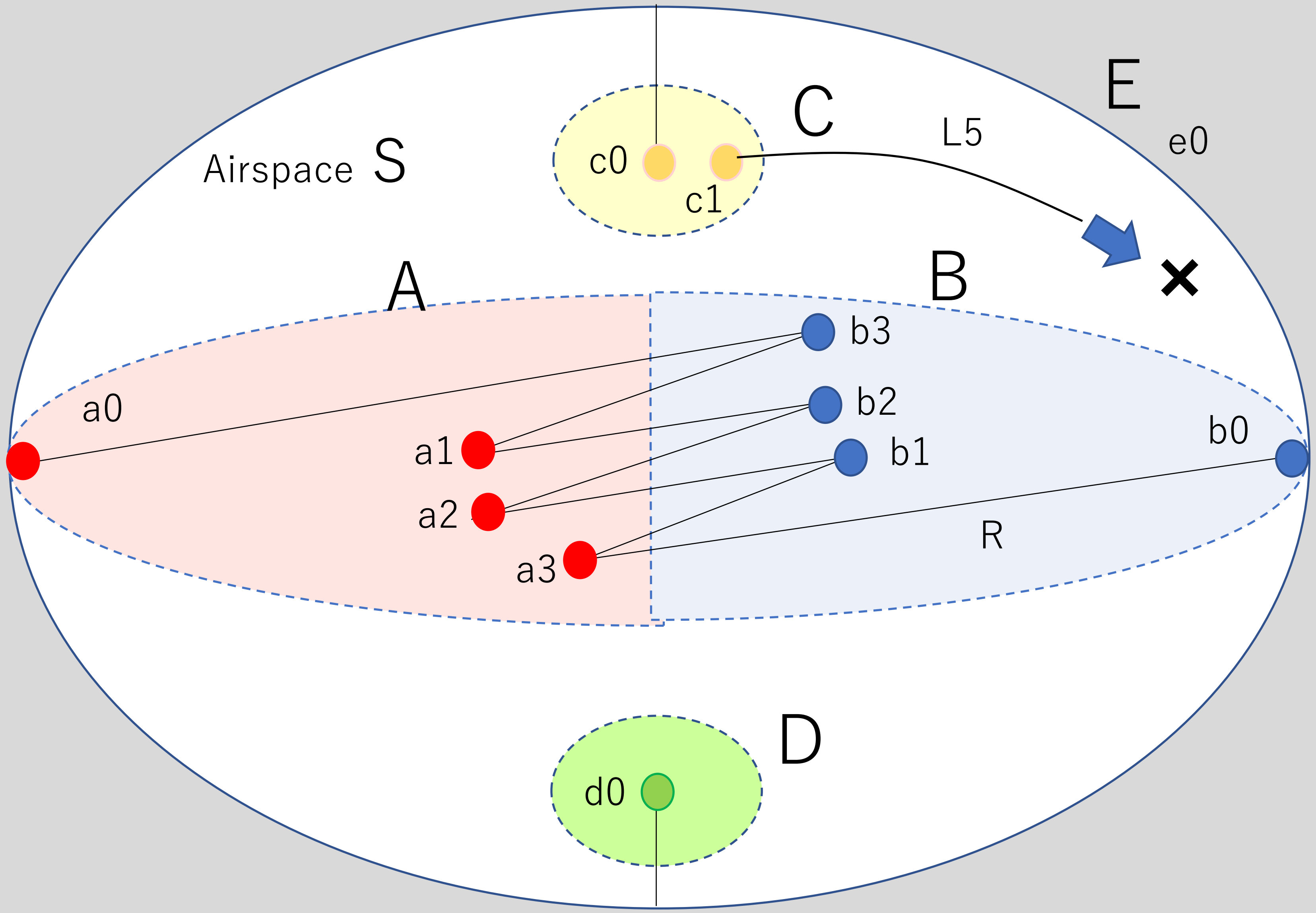


FIG.23

The color gamuts A and B are joined, and there is a Kempe chain R (drawn with a double line) in a0 to b0.
Both ends a0 and b0 of R are joined to e0 and there is no line r.
The line L5 from c1 in the color gamut C cannot be connected to a point belonging to the color gamut D.

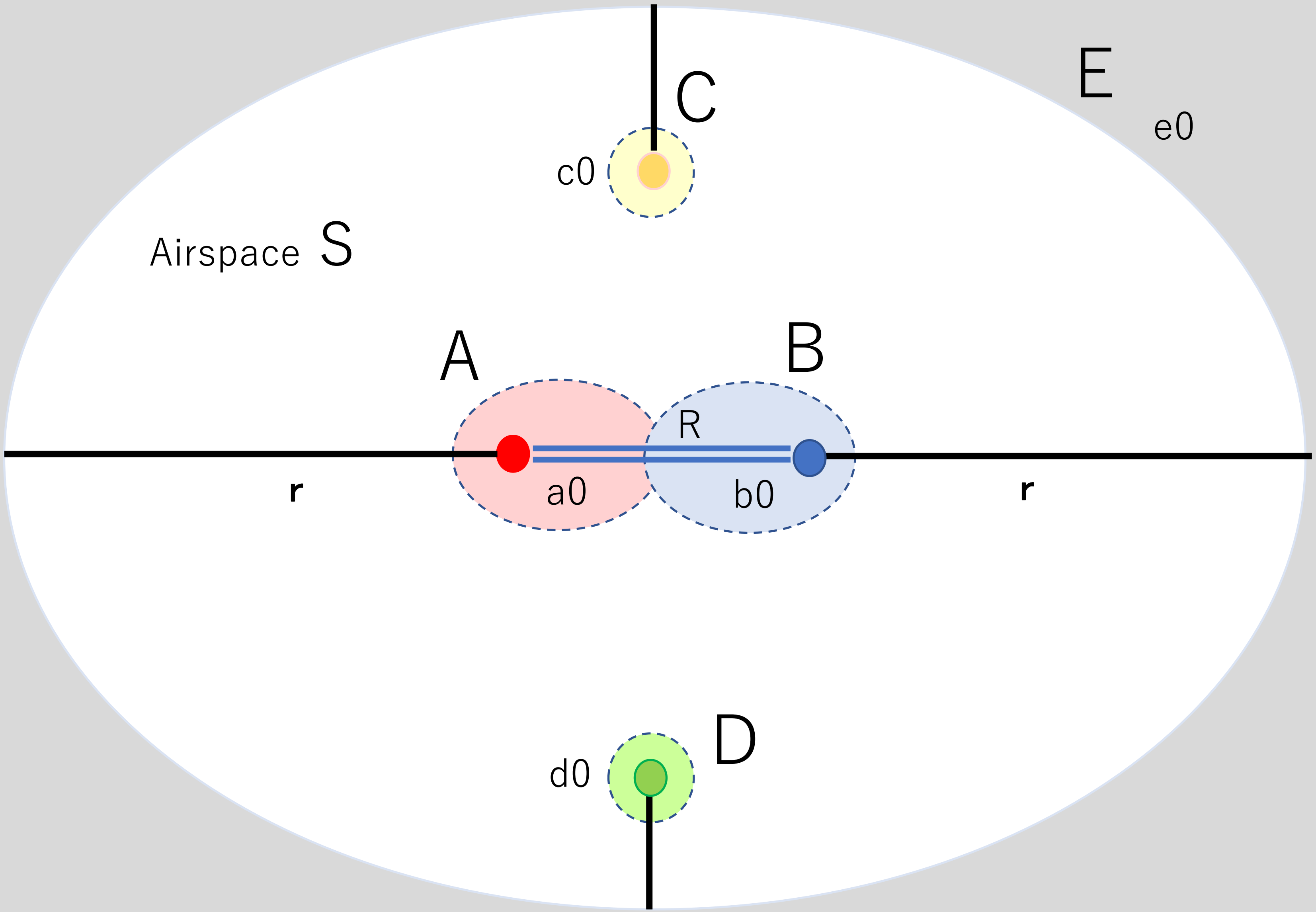


FIG.24

The line extending from e0 to a0, b0, c0, d0 is a straight rigid body.
The color gamuts A, B, C, and D are kept to a minimum.
Color gamuts A and B are joined.
There is a Kempe chain at a0 to b0.
The Kempe chain R is in the color gamuts A and B and does not appear in the airspace S.

Traditional proof of four-color problem

I used Google Translate from Japanese to English

June 28, 2021

https://4color-problem.net/