Ulmeanu Vlad   Crăciun Mihai   Hedeș Irina

Icleanu Ruxandra   Priboi Radu

coordonating teacher: Stoica Ioana

"Tudor Vianu" National High School of Computer Science

Bucharest

E L Y S I U M

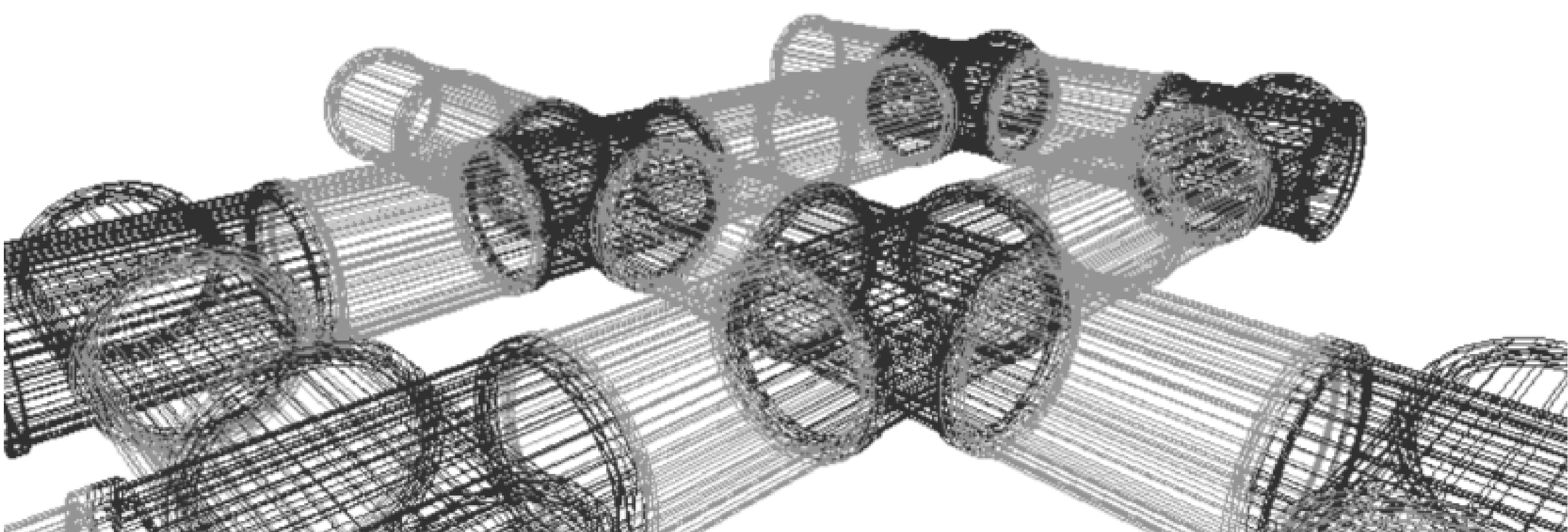


Table of contents

I.Introduction

The Space Settlement represents a very significant concern in the last half of the century. A very important question that arises is “Why should we consider moving past Earth?”. The greatest problem doesn’t lie within our planet, but in the fact that Earth is our only home. There have been enough times in Earth’s history when a major disaster has gravely affected the existing forms of life. We cannot know when the next catastrophe may happen, thus answering why we would want to consider leaving Earth behind. A Space Settlement provides an elegant way of creating a colony near another planet.

However, it is generally difficult to choose a consistent Settlement design regarding a great number of requirements, ranging from the cost and time needed for manufacturing the parts, to the habitants’ way of life. It is important to realize that each parameter’s requirements are complex to meet. Fortunately, we can exploit today’s immense computing power to create a model which satisfies all requirements. We want to achieve this by creating a great number of settlements starting from a number of randomized variables, and then rating them across all requirements, to lastly group them into categories using statistical methods.

This project aims to create a “perfect” settlement and place it into Mars’ vicinity, exploiting its basic resources to their fullest, while also solving some secondary issues that generally arise:

- diseases that humans are prone to developing in a 0g environment
- the existence of organisms that live in a rich CO_2 atmosphere on Mars that may present a threat to humans
- the existence of anaerobic organisms which arrive with humans on Mars and subsequently adapt to the atmosphere
- risks and possibility of an incident while collecting and producing breathable air and water

Why Mars?

One of the reasons for which we have chosen Mars is represented by the relatively small distance between it and the Earth, allowing this way to not only reach the planned spot for our settlement in a small amount of time, but as well send emergency resources from time to time.

Yet, we want to go to Mars because of the ease of acquiring raw materials. Choosing a location in the proximity of Mars means there is a constant flow of resources from the planet. For example, light metals as Magnesium, Aluminum, Titanium, Iron, and Chromium. These materials will be useful in constructions. Also, there have been found trace amounts of Lithium, Cobalt, Nickel, Copper, Zinc, Niobium, Molybdenum, Lanthanum, Europium, Tungsten and Gold. There is a real possibility that in some places these materials may be concentrated enough to be mined economically.

In addition, the Martian atmosphere consists of approximately 95% carbon dioxide, 2.7% nitrogen, 1.6% argon and traces of oxygen, carbon monoxide, water and methane, among other gases, resources that could ensure breathing.

Moreover, the Martian Geysers (or CO_2 jets) in the south polar region of Mars contain enough resources to sustain substantially more people than we can realistically send (200-1000 people), as we will demonstrate later.

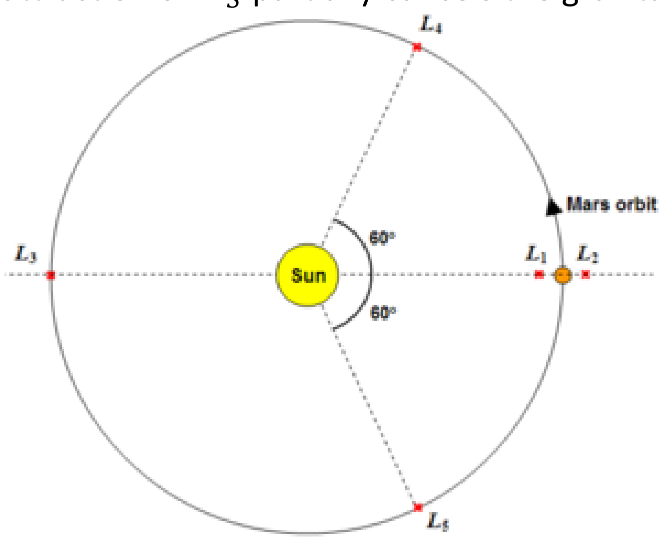
Mars, like Earth, has four seasons because the planet tilts on its axis, though the seasons vary in length. During the Martian summer, the polar ice cap, composed mainly of carbon dioxide ice, shrinks and may disappear. After that, when winter comes the ice cap grows back, thus creating the possibility of the existence of liquid water trapped beneath the carbon dioxide sheets. In 2004 it was confirmed that the southern polar cap has an average of 3 kilometers of thick slab of carbon dioxide ice with varying contents of frozen water, on average a mixture of 85% of CO_2 ice and 15% of water ice. It may not be enough water to sustain a settlement, but it would be a valuable contribution to our settlement's supplies.

II.Location and Positioning

The location of our settlement must be optimal considering various factors and possibilities. Some of the most important criteria in order to decide the placement were: the distance from Earth and the ease of acquiring raw materials, the existence of interplanetary dust and the stability of the placement.

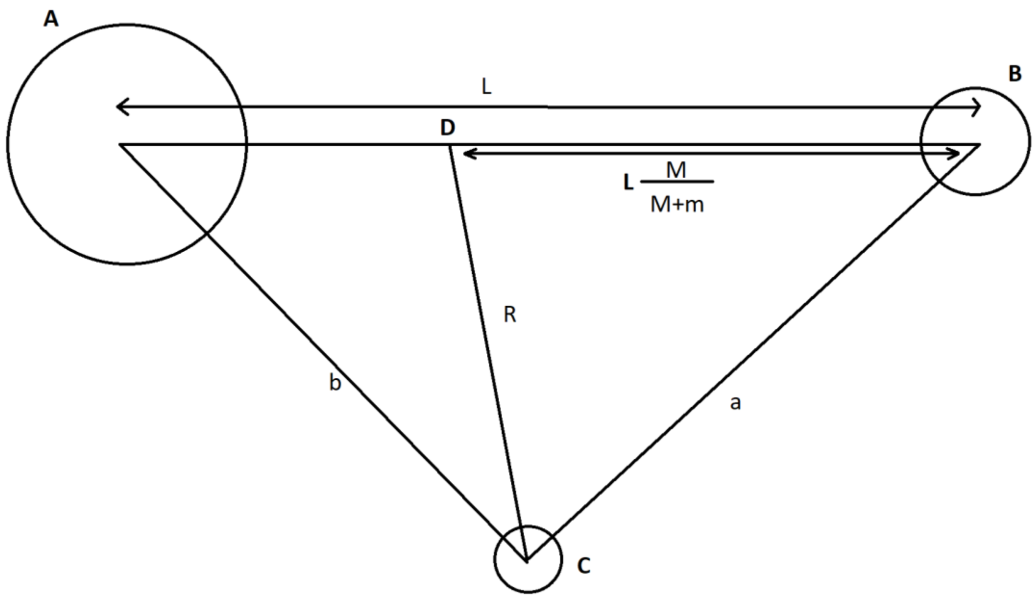
We have concluded that the L1 Lagrange point is one of the best one of the possible placements for our settlement. We have come to this conclusion by comparing L1 with the other five Lagrangian points.

Let M_S be the mass of the Sun, $M_M$the mass of Mars and m the mass of the spaceship. $k$is the Newton's constant of gravitation and L the distance between Sun and Mars. L1 lies on the line defined by the two large masses and is between them. The distance between the spacecraft and Mars is equal with $L \cdot \frac{M_S}{M_S+M_M}$.In this point the gravitational attraction of M_S partially cancels the gravitational attraction of M_M.



The benefits of placing the settlement in this location are significant. Firstly, the spacecraft orbiting about L1 can readily see the Sun and Earth, thus potentially simplifying spacecraft solar cell placement and the communication antenna design. Secondly, this location is the closest to the Earth, thus ensuring a minimum cost associated to placing the settlement in this position. And lastly, there are relatively small ΔV maneuvers required for the stationmaintenance.

The second Lagrange point is located on the same line, after the smaller of the two masses (in our case Mars). The L3 point is on the same line, but beyond the bigger one (in our system represented by the Sun). In both cases, the spacecraft would be too far away from the Earth, so these placements do not represent realistic options.
The L4 and L5 are positioned so as to form an equilateral triangle with the two larger masses, the Sun and Mars. Let us present a proof of this in the following:



If A is the position of Sun, B is the position of Mars and C is the spacecraft, the spacecraft, which is in equilibrium with the Sun and Mars will always keep the same distance from them. Let the center of rotation be point D and all three bodies have the same orbital period$T$. As C is motionless in the rotating frame, there is no Corriolis force. The spacecraft will sense a centrifugal force and the Sun and Mars will too.

Let R be the rotation radius of the spacecraft, V the velocity of Mars and v the velocity of the spacecraft. Then, the rotation radius of Mars is $L \cdot \frac{M_S}{M_S+M_M}$.

Because $distance = velocity \cdot time$, we have :
$2 \cdot \pi \cdot R = v \cdot T$ and $2 \cdot \pi \cdot L \cdot \frac{M_S}{M_S+M_M} = V \cdot T$

$$\frac{2\pi}{T} = \frac{v}{R}$$

$$\frac{2\pi}{T} = \frac{V}{L} \cdot (1 + \frac{M_M}{M_S})$$

Therefore $\frac{v}{R} = \frac{V}{L} \cdot (1 + \frac{M_M}{M_S})(1)$

The centrifugal force on Mars is

$$F_{cM} = \frac{M_M \cdot V^2}{L \cdot \frac{M_S}{M_S + M_M}} = \frac{M_M \cdot V^2 \cdot (1 + \frac{M_M}{M_S})}{L}$$

And it is balanced by the pull of the Sun

$$F_a = \frac{k \cdot M_S \cdot M_M}{L^2}$$

Thus, we have $\frac{k \cdot M_S \cdot M_M}{L^2} = \frac{M_M \cdot V^2 \cdot (1+\frac{M_M}{M_S})}{L}$ and this is equivalent to $\frac{k \cdot M_S}{L} = V^2 \cdot (1 + \frac{M_M}{M_S})(2)$

The centrifugal force on the spacecraft is:

$$F_{a1} = \frac{m \cdot v^2}{R}$$

This force must be balanced by the attracting forces of the Sun and of Mars. Only the the components along the line R are effective in balancing the centrifugal force. Let α and β be the two angles in which R divides the angle C.

The attracting forces of the Sun and of Mars are $\frac{k \cdot m \cdot M_S}{b^2}$ and $\frac{k \cdot m \cdot M_M}{a^2}$ respectively.

Therefore, we obtain:

$\frac{v^2}{R} = \frac{k \cdot M_M}{a^2} \cdot \cos\beta + \frac{k \cdot M_S}{b^2} \cdot \cos\alpha (3);$

The forces pulling the spacecraft in directions perpendicular to R must cancel so we obtain $\frac{k \cdot m \cdot M_S}{b^2} \cdot \sin\alpha = \frac{k \cdot m \cdot M_M}{a^2} \cdot \sin\beta \Leftrightarrow \frac{M_M}{a^2} \sin\beta = \frac{M_S}{b^2} \sin\alpha (4)$

By squaring both sides of equation (1) and then multiplying by L^2 and then dividing by $(1 + \frac{M_M}{M_S})$ we have :

$$\frac{v^2 \cdot (\frac{L^2}{R^2})}{1 + \frac{M_M}{M_S}} = V^2 \cdot (1 + \frac{M_M}{M_S})(5)$$

Using the relation (2) in (5) gives:

$$\frac{v^2 \cdot (\frac{L^2}{R^2})}{1 + \frac{M_M}{M_S}} = \frac{k \cdot M_S}{L}$$

Next we multiply both sides by $(1 + \frac{M_M}{M_S})$, divide by L^2 and multiplyby R:

$$\frac{v^2}{R} = \frac{k \cdot M_S}{L^3} \cdot R \cdot (1 + \frac{M_M}{M_S})$$

Using relation (3) we obtain that :

$$\frac{k \cdot M_S}{L^3} \cdot R \cdot (1 + \frac{M_M}{M_S}) = \frac{k \cdot M_M}{a^2} \cdot \cos\beta + \frac{k \cdot M_S}{b^2} \cdot \cos\alpha$$

We divide everything by $k \cdot M_S$ and we have :

$\frac{1}{L^2} \cdot \frac{R}{L} \cdot (1 + \frac{M_M}{M_S}) = \frac{1}{b^2}\cos\alpha + \frac{M_M}{M_S}\cos\beta$ (6);

From now on we suppose that triangle ABC is equilateral, thus $a = b = L = r$ and all of its angles have the value of 60 degrees.

We multiply equations (6) and (4) by r^2:

$\frac{R}{c}(1 + \frac{M_M}{M_S}) = \frac{M_M}{M_S}\cos\beta + \cos\alpha$ (7)

$$M_M \sin\beta = M_S \sin\alpha \Leftrightarrow \frac{M_M}{M_S} = \frac{\sin\alpha}{\sin\beta} \, (8)$$

We substitute this relation on the right side of (7) and we obtain:

$\frac{R}{c}(1 + \frac{M_M}{M_S}) = (\frac{\sin\alpha\cos\beta}{\sin\beta}) + \cos\alpha;$

If we multiply by $\sin\beta$ we get $\sin\beta \frac{R}{c}(1 + \frac{M_M}{M_S}) = \sin\alpha\cos\beta + \sin\beta\cos\alpha = \sin(\alpha + \beta) = \sin C = \sin B$

Then we divide both sides by R and we have:

$$\frac{\sin\beta}{\frac{LM_S}{M_S + M_M}} = \frac{\sin B}{R}$$

This relation is the law of sines in the triangle CBD meaning that our prediction that the triangle ABC is equilateral is true.

We consider that L4 and L5 are the most stable of the five Lagrangian points. Even if there are perturbations that interfere with the body placed in one such point, the action of the Corriolis force will make it drift back toward its initial position. This makes these two points the best decision we can make for the placement taking in consideration only the stability of the location. But, by analyzing this locations more, we observed that the distance from Mars to either of these two Lagrange points is the same as the distance from Mars to the Sun ($\cong 228 \cdot 10^6 km$). This distance makes communication with the settlement unrealistic as the power and size of the needed equipment would be on the

order of the Goldstone Deep Space Network facility in California.
Also, while L4 and L5's stability can be exploited, this same stability also attracts a multitude of interplanetary body, making these placements fairly dangerous. From our point of view, the benefits of inherent stability of these locations do not outweigh the risks of residing our settlement here.

III.Designing the Space Settlement

The Space Settlement needs to be perfect in an enormous amount of ways, including:

- *(1)*The cost and time needed for manufacturing the parts
- *(2)*The cost for assembling the parts on an Earth orbit
- *(3)*The complexity of the part assembly
    - How much time and effort are needed?
    - How important are the risks?
- *(4)*The settlement capacity
- *(5)*How easy it is to increase the settlement's capacity instead of limiting through other methods the number of occupants
- The habitants' way of life, which can be described through:
    - *(6)*Available surface for each inhabitant
        - We obviously don't want the surface to be small. However, if we allocate too much space for one human, this simply means that more people could be sent on the Settlement, facilitating many operations.
    - *(7)*The average distance between two residences
        - The higher, the harder it will be for two inhabitants to spend time together
        - If it's too low, other problems appear, including less personal freedom, as well as an increased chance of the emergence of diseases or epidemics
    - *(8)*The similarities the Settlement has compared to Earth
        - Could it have a sky? (toruses vs. modular parts)
        - How much volume of the settlement has artificial gravity?
    - The risks that each inhabitant faces during their everyday routine, both:
        - *(9)*On the short run (during work)

- (10)On the long run (mental or physical diseases)
- *(11)*How often and in what quantity do we need to resupply the Settlement?
- *(12)*How easy, or difficult it is to repair a part of a Settlement if a disaster occurs
  - How probable would it be for a disaster to occur?

Observation:

- Including modular parts will have a great impact on overall cost, as well as the complexity of the assembly, this being the reason why we have chosen them.

PCA (Principal Component Analysis)

We want to analyze a considerable number of Settlements that we created using our 12 remarks.

There are multiple questions that arise when trying to figure out the design of a good Settlement, based on the remarks that we have previously listed:

- How could we know which remarks are the most meaningful when compared to each other?
- How can we find out which remarks are related, and to what degree?
- How can we visually group a number of Settlements in multiple clusters, in relation with some of their prominent features?

In order to answer these questions, we have chosen to perform a Principal Component Analysis on our test data, considering each settlement as a variable, each of them having multiple factors, that are scores (from 1 to 10) related with the previous remarks.

We can apply PCA on a chart of data (A), with variables on rows and factors on columns.

By studying the variables (Settlements) we can:

- Say when two variables are similar or not
- If it is possible to group the variables in typologies

By studying the factors (Scores in tests) we can:

- Interpret similarities between variables

- Talk about relations between the factors
    - PCA only analyzes linear relations, from which we can obtain a correlation coefficient for each pair of variables

The Correlation Matrix

If our data matrix *A* has *n* Settlements and *m* variables, then our Correlation Matrix *ρ*has*m* lines, and *m* columns.

- ***ρ***$\boldsymbol{\rho_{i,j}}$represents the coefficient of correlation between the variables***i***and***j****.*

The coefficient of correlation is always located between -1 and 1.

- As it gets closer to 1 in absolute, the two variables are more and more related, thus needing the other one less in our calculations.
- If the coefficient is nearing 0, it means that the two variables aren't correlated, resulting in needing both in our calculations.

The Correlation Matrix is normally used when different variables aren't measured on the same scale, or they can't be measured in the same way:

- If a variable would represent the time needed to deploy the Settlement, then lower would be better, but we would desire a bigger settlement capacity. This means that we cannot justify regarding a lower or bigger value as the best.

We have used the *Pearson Correlation Coefficient* in order to calculate the matrix *ρ.*

- We define x_i as the vector corresponding to the column of data for the ith variable
- We also define the $\overline{x_i}$ as the mean value of the parameters found in the x_i vector.

Then our correlation coefficient is:

$$\rho_{i,j} = \frac{\sum_{k=1}^{m}\left(x_{i,k} - \overline{x_i}\right) \cdot \left(x_{j,k} - \overline{x_j}\right)}{\sqrt{\sum_{k=1}^{m}(x_{i,k} - \overline{x_i})^2} \cdot \sqrt{\sum_{k=1}^{m}(x_{j,k} - \overline{x_j})^2}}$$

Observations:

- $\rho_{i,i} = 1 \; \forall \; i = \overline{1,m}$, meaning that if we would know all the factors from any vector, receiving it again wouldn't help us at all, since it has exactly the same measurements.
    - The previously listed formula verifies the made assumption.

- $\rho_{i,j} = \rho_{j,i} \forall \; i = \overline{1,m}$, $\forall \; j = \overline{1,m}, i \neq j$, meaning that we would know as much about the other vector, either by only having x_i or x_j.
    - This means that the correlation matrix is symmetric.

- Since ρ is symmetric, and $\rho_{i,j} \epsilon \mathbb{R} \; \forall \; i = \overline{1,m}$, $\forall \; j = \overline{1,m}$, then $\lambda_i \epsilon \mathbb{R} \; \forall \; i = \overline{1,m}$.
  Proof:
  We consider $\bar{x}$ to be the conjugated value of x.
    - Let λ be an eigenvalue of ρ and V the eigenvector corresponding to λ.
    - As ρ is symmetric, we deduce that $\rho = \rho^t$. Also,as $\rho_{i,j} \epsilon \mathbb{R}$, we know that $\rho = \bar{\rho}$.
    - Therefore, we have that $\lambda \bar{V}^t V = \bar{V}^t \lambda V = \bar{V}^t \rho V = \bar{V}^t \bar{\rho}^t V = (\bar{\rho}\bar{V})^t V = \bar{\lambda}\overline{V^t} V$.
    - This means that $\lambda = \bar{\lambda}$, that is equivalent to $\lambda \epsilon \mathbb{R}$.

- Since ρ is a correlation matrix, and all its parameters are real, then ρ is positive semidefinite, meaning that $\lambda_i \geq 0 \forall \; i = \overline{1,m}$.

We want to calculate the eigenpairs, that are formed from eigenvalues($\lambda_i, i = \overline{1,m}$) and their correspondent eigenvector($V_i, i = \overline{1,m}$).

- All eigenpairs satisfy the following equations:
    - $\rho V_i = \lambda_i V_i$
    - $\det(\rho - \lambda_i I_m) = 0$
- We define $||V||_F$ as the Frobenius norm of eigenvector V.
    - $||V||_F = \sqrt{\sum_{k=1}^{m} x_k^2}$, where $V = \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix}$.

- If V is any eigenvector that belongs to ρ, then$a \cdot V$, where a is a scalar is also an eigenvector that belongs to ρ.
- Therefore, $\frac{1}{||V||_F} V$ is also an eigenvector that belongs to ρ.
- We will choose only the eigenvectors that belong to ρ and have the Frobenius norm, or modulus equal to 1. Let be S the set of the eigenvectors of matrix ρ. Also, we define the set $M = \{ V' = \frac{1}{||V||_F} V |\ V \epsilon S \}$.

We have settled on the *Power Method* algorithm, which needs:

- $\rho$to have m linearly independent eigenvectors.
- The eigenvalues to be arranged in such a way that $|\lambda_1| > |\lambda_2| > \cdots > |\lambda_m|$, even if we only need $|\lambda_1| > |\lambda_2| \geq \cdots \geq |\lambda_m|$ initially, since ρ has to have a dominant eigenvalue.

The Power Method algorithm begins with a randomized vector, that has m lines: B_1

Afterwards, it defines the following recurrent relation:

$$B_{i+1} = \frac{\rho B_i}{||\rho B_i||_F}$$

It can be demonstrated that $\lim_{n \to \infty} B_n = V_1$, thus giving us the eigenvector that is paired with the dominant eigenvalue.

- For a large enough n we would have both $B_n \to V_1$, and $B_{n+1} \to V_1$.
- By replacing in the recurrent relation, we obtain:

$$V_1 \cong \frac{\rho V_1}{||\rho V_1||_F}$$

However, we know that $\rho V_1 = \lambda_1 V_1$. $||\rho V_1||_F = ||\lambda_1 V_1||_F = \sqrt{\lambda_1^2 (x_1^2 + x_2^2 + \cdots + x_n^2)} = \lambda_1 ||V_1||_F$

By replacing again, we have:

$$V_1 \cong \frac{\lambda_1 V_1}{\lambda_1 ||V_1||_F}$$

- In order to obtain the wanted equality, we would need $||V_1||_F$ to be equal to 1, which we already know is true.

$$(\rho - \lambda_1 I_m)V_1 = O_m \Leftrightarrow \begin{pmatrix} 1-\lambda_1 & \rho_{1,2} & & \\ \rho_{2,1} & 1-\lambda_1 & & \\ & & \ddots & \\ & & & 1-\lambda_1 \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix} =$$

$$= \begin{pmatrix} (1-\lambda_1)x_1 + \rho_{1,2}x_2 + \cdots + \rho_{1,m}x_m \\ \vdots \\ \end{pmatrix} = O_m$$

We have obtained that $(1-\lambda_1)x_1 + \rho_{1,2}x_2 + \cdots + \rho_{1,m}x_m = 0$, and from here we get that the dominant eigenvalue:

$$\lambda_1 = 1 + \frac{\rho_{1,2}x_2 + \cdots + \rho_{1,m}x_m}{x_1}$$

In practice, after at most 100 iterations, we can correctly estimate λ_1's first 8 decimals.

- We define the matrix: $U_1 = \frac{V_1}{||V_1||_F}$

It is known that the matrix $\rho_2 = \rho - \lambda_1 U_1 U_1^t$ has $0, \lambda_2, \lambda_3, \ldots, \lambda_m$ as its eigenvalues, with λ_2 obviously being its dominant eigenvalue, since all of them are positive.

- We can apply the same *Power Method* algorithm again for the ρ_2 matrix, obtaining both $V_2$and λ_2.
- By using the *Power Method* algorithm for another $m-2$ times, we can obtain all of the eigenvectors, as well as their correspondent eigenvalues.

The matrix that is obtained by merging the eigenvectors is named V/.

By performing the dot product of the A (data matrix) and V/, we obtain the *Score Matrix,* which has m *Principal Component* vectors.

Example

Suppose that a Space Agency is looking for an athlete, that has to be good from multiple points of view. The easiest way to find one is to look at some recent results from a decathlon contest. These are the results from the 2018 Commonwealth games:

↓	C1-T	C2	C3	C4	C5	C6	C7	C8	C9	C10
	athletes	100m	long jump	shot put	high jump	400m	110m hurdles	discus throw	pole vault	javelin throw
1	Lindon Victor	10,70	7,24	15,79	2,01	49,48	14,87	52,32	4,6	71,10
2	Kyle Cranston	11,16	7,18	13,59	1,92	49,94	15,12	43,19	4,4	62,36
3	Kurt Felix	11,20	7,26	15,24	1,95	50,49	15,25	48,04	4,2	67,47
4	Atsu Nyamadi	11,27	7,25	14,49	2,01	50,45	14,82	46,61	3,3	54,63
5	Curtis Mathews	11,39	6,89	12,34	1,89	50,71	15,63	40,59	3,9	49,86
6	Ben Gregory	11,60	6,94	12,80	1,89	50,31	15,16	38,85	4,8	57,30

(Fig.1)

To easily understand the results, we have only included 6 athletes and 9 trials.

Observations:

- It is very important to have all athletes finish all the trials, since we cannot construct a partial correlation matrix.
    - Having a partial correlation matrix can lead to other unwanted problems, such as negative eigenvalues, that would break our Power Method algorithm.
- We can observe that the unit of measurement differs, from seconds to meters. However, this doesn't affect us, since all the athletes score in the same unit of measurement in any trial.
- We can also recognize that there isn't one direction in which the scores increase. For example, a lower value in the 100m trial would be desired, while a larger one in the long jump trial is better.
- While in our example the correlation matrix would have 9 lines and 9 columns, calculating it by hand wouldn't be only a chore, but also error-prone. Also, since our actual data matrix is much larger, we would want to write a code that can automate the process.

We have decided to write the code in C++. We have verified our partial and final results using both Minitab and Wolfram Mathematica.

```cpp
inline double sq ( double x )
{
  return x * x;
}
```

```cpp
fin >> n >> m; /// n variables, m factors
```

```cpp
128 int i, j, k;
129 for ( i = 0; i < n; i++ )
130   for ( j = 0; j < m; j++ )
131     fin >> data[i][j];
132
133 for ( j = 0; j < m; j++ )
134 {
135   for ( i = 0; i < n; i++ )
136     avg[j] += data[i][j];
137   avg[j] /= n;
138 }
```

```cpp
140 double ss, ss1, ss2;
141 for ( i = 0; i < m; i++ )
142   for ( j = 0; j < m; j++ )
143   {
144     ss = ss1 = ss2 = 0.0;
145     for ( k = 0; k < n; k++ )
146     {
147       ss += ( data[k][i] - avg[i] ) * ( data[k][j] - avg[j] );
148       ss1 += sq(data[k][i]-avg[i]);
149       ss2 += sq(data[k][j]-avg[j]);
150     }
151     ss1 = sqrt ( ss1 );
152     ss2 = sqrt ( ss2 );
153
154     if ( ss1 != 0 && ss2 != 0 )
155       mcorel[i][j] = ss / ss1 / ss2;
156     else
157       mcorel[i][j] = inf;
158   }
```

Observations (Fig.2):

- We read and store the data matrix on lines 128-131.
- We calculate the average values of each column vector (⇔each factor) over all of their entries (⇔variables) on lines 133-138.
- We calculate the 3 sums that are required by the Pearson Correlation Coefficient on lines 147-149.
- We have to make sure that we can calculate the correlation coefficient by having a nonzero denominator on line 154.

```
Correlation[A]

{{1., -0.677479, -0.802278, -0.719246, 0.782983, 0.488038, -0.898415, -0.108543, -0.753909},
 {-0.677479, 1., 0.896905, 0.826491, -0.447742, -0.720641, 0.841776, -0.212694, 0.689137},
 {-0.802278, 0.896905, 1., 0.856097, -0.552052, -0.66682, 0.965973, -0.00785529, 0.829372},
 {-0.719246, 0.826491, 0.856097, 1., -0.460794, -0.811285, 0.88447, -0.363087, 0.478347},
 {0.782983, -0.447742, -0.552052, -0.460794, 1., 0.616443, -0.573085, -0.539761, -0.742836},
 {0.488038, -0.720641, -0.66682, -0.811285, 0.616443, 1., -0.599694, 0.0518492, -0.470793},
 {-0.898415, 0.841776, 0.965973, 0.88447, -0.573085, -0.599694, 1., -0.0921846, 0.757187},
 {-0.108543, -0.212694, -0.00785529, -0.363087, -0.539761, 0.0518492, -0.0921846, 1., 0.514795},
 {-0.753909, 0.689137, 0.829372, 0.478347, -0.742836, -0.470793, 0.757187, 0.514795, 1.}}
```

1.0000	-0.6775	-0.8023	-0.7192	0.7830	0.4880	-0.8984	-0.1085	-0.7539
-0.6775	1.0000	0.8969	0.8265	-0.4477	-0.7206	0.8418	-0.2127	0.6891
-0.8023	0.8969	1.0000	0.8561	-0.5521	-0.6668	0.9660	-0.0079	0.8294
-0.7192	0.8265	0.8561	1.0000	-0.4608	-0.8113	0.8845	-0.3631	0.4783
0.7830	-0.4477	-0.5521	-0.4608	1.0000	0.6164	-0.5731	-0.5398	-0.7428
0.4880	-0.7206	-0.6668	-0.8113	0.6164	1.0000	-0.5997	0.0518	-0.4708
-0.8984	0.8418	0.9660	0.8845	-0.5731	-0.5997	1.0000	-0.0922	0.7572
-0.1085	-0.2127	-0.0079	-0.3631	-0.5398	0.0518	-0.0922	1.0000	0.5148
-0.7539	0.6891	0.8294	0.4783	-0.7428	-0.4708	0.7572	0.5148	1.0000

Afterwards, we wish to calculate the eigenpairs using the Power Method.

```
73  void poweriteration ( double orig[][maxn+5], int m, int nr_it )
74  {
75      int i, j;
76      double mt[maxn+5][maxn+5];
77      for ( i = 0; i < m; i++ )
78          for ( j = 0; j < m; j++ )
79              mt[i][j] = orig[i][j];
80      double v[maxn+5], u[maxn+5];
81      for ( i = 0; i < m; i++ )
82          v[i] = rand() % INT_MAX;
83  
84      double norm;
85      while( nr_it-- )
86      {
87          dotproductMxV ( mt, v, m, u );
88          for ( i = 0, norm = 0.0; i < m; i++ )
89              norm += sq(u[i]);
90  
91          for ( i = 0, norm = sqrt ( norm ); i < m; i++ )
92              v[i] = u[i] / norm;
93      }
94  
95      double ev, s = 0.0;
96      for ( i = 1; i < m; i++ )
97          s += mcorel[0][i] * v[i];
98      ev = 1.0 - s / (-v[0]);
99  
100     for ( i = 0; i < m; i++ )
101         eigenvectors[szev][i] = v[i];
102     eigenvalues[szev++] = ev;
103 }
```

Observations (Fig.3) :

- The "*poweriteration*" function receives the square matrix from which we want to extract the dominant eigenpair. It also receives the number of rows/columns (*m*), and the number of iterations we want to do (*nr_it*) – Line 73
- The candidate eigenvector starts with *m* randomized values – Line 82
- We calculate on line 87 the dot product between the matrix we received, and the candidate eigenvector. We store the result in a column vector named *u*.
- On lines 88-91 we calculate the Frobenius norm of the *u* column vector.
- We then actualize the candidate vector on line 92 by dividing the dot product with the calculated norm.

- We apply the previously mentioned formula to obtain the dominant eigenvalue, knowing that the error between the candidate eigenvector and the actual one is insignificant – Lines 96-98

Now we need to transform our current matrix into another one with the same eigenvalues, except the dominant one (which we already know), which is replaced by 0.

- It is notable that our current matrix has the same eigenvectors as the next matrix, even if it has a different eigenvalue.
- We need to calculate $U_k = \frac{V_k}{||V_k||_F}, k = \overline{1, m}$ (*k* represents how many times we have called the *poweriteration* function). However, since the eigenvectors don't change as the matrixes change their values (they originally belonged to the correlation matrix), we know that their Frobenius norm is 1.
  - Thus, $U_k = V_k$. This is an useful shortcut that we have used in our code.

```
169   for ( i = 0; i < m; i++ )
170   {
171     poweriteration( cm, m, 100 );
172     for ( j = 0; j < m; j++ )
173       uu[j][0] = uut[0][j] = eigenvectors[i][j];
174     /// U*(U^T) == V*(V^T)
175     dotproduct( uu, {m,1}, uut, {1,m}, cm2 );
176
177     for ( j = 0; j < m; j++ )
178       for ( k = 0; k < m; k++ )
179         cm[j][k] -= eigenvalues[i] * cm2[j][k];
180   }
```

Observations (Fig.4):

- "*cm*" is an auxiliary matrix that starts as a copy of the correlation matrix
- "*uu*" is the column vector U_k, and "*uut*" is the line vector U_k^t
- At lines 172-173, we copy the eigenvector that we have just calculated in the power iteration function in "*uu*", and "*uut*".
- On line 175 we compute the dot product between U_k and U_k^t in "*cm2*", another auxiliary matrix.

- Finally, we subtract $\lambda_k U_k U_k^t$ from the matrix we had at the beginning of the loop, obtaining the one we need for the following cycle (in Fig.4 the cycle index is marked as *i*)

```
#define pii pair<int,int>
```

```
void dotproduct ( double mt[][maxn+5], pii d1, double mt2[][maxn+5], pii d2, double mt3[][maxn+5] )
{
  int i, j, k, n = d1.first, m = d2.second;
  for ( i = 0; i < n; i++ )
    for ( j = 0; j < m; j++ )
      for ( mt3[i][j] = 0, k = 0; k < d1.second; k++ )
        mt3[i][j] += mt[i][k] * mt2[k][j];
}
```

```
182    transpose ( eigenvectors, {m,m} );
183    dotproduct ( data, {n,m}, eigenvectors, {m,m}, scores );
```

Observations (Fig.5):

- Our dot product function takes in two matrixes “*mt*”, and “*mt2*”, and saves the result in “*mt3*”.
    - The dimensions of “*mt*” and “*mt2*” can be found in “*d1*” and “*d2*”.
- We have to transpose our eigenvectors on line 182 because we have stored them as line vectors, and we need them as column vectors for the score matrix.
- The score matrix is obtained by doing a dot product between the initial “data” matrix and the V matrix, containing the eigenvectors.

```
Eigenvalues:
5.99619565
1.79431331
0.64289995
0.42418490
0.14240619
0.00000000
0.00000000
0.00000000
0.00000000
Eigenvectors:
-0.3629 0.3622 0.3912 0.3576 -0.3018 -0.3136 0.3885 0.0230 0.3396
0.1019 0.2072 0.0671 0.3211 0.4042 -0.1171 0.1000 -0.7319 -0.3412
-0.2580 0.0542 0.2192 -0.1597 0.3466 0.7594 0.3001 -0.0777 0.2532
0.5809 0.4085 0.2604 -0.1519 0.4138 -0.1805 -0.1213 0.2222 0.3699
0.1980 -0.6687 0.3196 0.3944 0.2777 -0.0917 0.3003 0.2771 -0.0588
0.1979 -0.6685 0.3196 0.3949 0.2775 -0.0918 0.3001 0.2772 -0.0588
0.1979 -0.6685 0.3196 0.3949 0.2775 -0.0918 0.3001 0.2772 -0.0588
0.1979 -0.6685 0.3196 0.3949 0.2775 -0.0918 0.3001 0.2772 -0.0588
0.1979 -0.6685 0.3196 0.3949 0.2775 -0.0918 0.3001 0.2772 -0.0588
Score matrix:
30.6166 22.7982 26.8771 21.7866 16.4513 18.7124
0.1609 2.3909 1.6269 6.5312 6.8827 3.4679
62.5601 57.3821 60.7873 56.6436 53.7852 54.6311
51.7458 49.4048 51.3473 46.4648 44.9064 48.2509
28.2936 25.5086 27.2431 27.1529 25.3203 24.6965
28.2791 25.4950 27.2288 27.1384 25.3063 24.6833
28.2791 25.4950 27.2288 27.1384 25.3063 24.6833
28.2791 25.4950 27.2288 27.1384 25.3063 24.6833
28.2791 25.4950 27.2288 27.1384 25.3063 24.6833
```

Eigenanalysis of the Correlation Matrix

Eigenvalue	5,9962	1,7943	0,6429	0,4242	0,1424	0,0000	-0,0000	-0,0000	-0,0000
Proportion	0,666	0,199	0,071	0,047	0,016	0,000	-0,000	-0,000	-0,000
Cumulative	0,666	0,866	0,937	0,984	1,000	1,000	1,000	1,000	1,000

Eigenvectors

Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9
100m	-0,363	0,102	-0,258	0,581	0,198	-0,052	-0,300	-0,553	0,126
long jump	0,362	0,207	0,054	0,409	-0,669	0,240	0,047	-0,163	-0,350
shot put	0,391	0,067	0,219	0,260	0,320	-0,714	0,208	-0,064	-0,259
high jump	0,358	0,321	-0,160	-0,152	0,394	0,421	0,471	-0,384	0,135
400m	-0,302	0,404	0,347	0,414	0,278	0,286	0,159	0,512	-0,088
110m hurdles	-0,314	-0,117	0,759	-0,180	-0,092	0,045	0,161	-0,491	-0,030
discus throw	0,389	0,100	0,300	-0,121	0,300	0,235	-0,755	-0,038	-0,131
pole vault	0,023	-0,732	-0,078	0,222	0,277	0,329	0,139	0,010	-0,452
javelin throw	0,340	-0,341	0,253	0,370	-0,059	0,074	0,037	0,105	0,739

(Fig.6)

Observations (Fig.6):

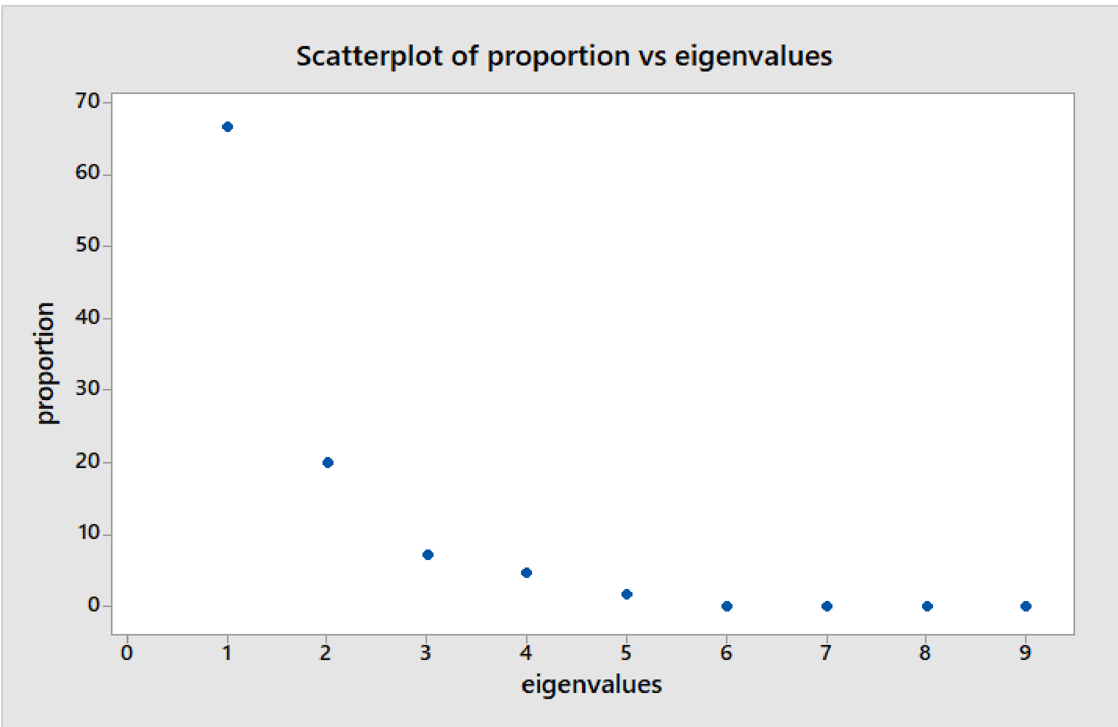
- Our eigenvalues are identical to the ones we obtained in Minitab, with the exception of the last 4 eigenvectors. This happened because the Power Method algorithm replaced all nonzero eigenvalues with zeroes along each cycle.
    - Therefore, in the last four cycles all eigenvalues of the current matrix were either 0, or very close to 0. Thus, the current matrix, and as a consequence the calculated eigenvectors didn't vary.
    - As we said earlier, the Power Method algorithm requires us to be able to sort the eigenvalues in a way such that we obtain $|\lambda_1| > |\lambda_2| > \cdots > |\lambda_m|$. However, most of the times we cannot assure this claim. However, the closer in absolute value the eigenvalue is to 0, the more insignificant it becomes. Fortunately, those last incalculable eigenvectors are useless for our analysis, since of their very small proportion relative to $Tr(\rho)$.

Interpretation

By interpreting PC1 we can find out that:

- The scores in "*discus throw*" (0,389) and "*shot put*" (0,391) are strongly related, being relatively high and close to each other.
- The scores for "*long jump*", "*shot put*", "*high jump*", "*discus throw*", and "*javelin throw*" were related.
- The scores for "*100m*", "*400m*", and "*110m hurdles*" are related.
    - They aren't related with the previously mentioned scores, because of their different sign, and large absolute value. This proves to us that the Correlation Matrix realizes that lower is better for "*100m*" and the ones related to it, while larger is better for the other group.
- The score for "*pole vault*" is unrelated. It having its absolute value close to 0 shows that its scores can hardly be put in relation with any other trial.

- By plotting the eigenvalues in the order that they were found by the Power Method algorithm (after dividing it by $Tr(\rho) = \sum_{k=1}^{m} \lambda_k = m$), we can come to the conclusion that we only need two factors out of the 9 (*86,6%*) to perform a correct analysis, and we can hardly benefit (*98,4%*) from including another two.



Settlement Generator

We wish to quickly generate a large number of Space Settlements, and later analyze them using the Principal Component analysis, taking into consideration multiple criteria.

- The generator needs to use multiple pieces in order to construct various settlements. The ones we have designed are listed below:

The designs of each container can be found in the "*Room Management*" part below.

- **(0)**4-way connector
  - $l_1 = l_2 = 15\ m, r = 5\ m$
- 2-way standard container:
  - $l = 30\ m, r = 20\ m$
- **(1)**Living residence: $n = 5\ layers$
- **(2)**School/Research area
- **(3)**Agricultural area: $n = 10\ layers$
- **(4)**Nuclear energy area
- **(5)**Hospital area
- **(6)**Air/water refinery
- **(7)**Solar panels

Allocated space per human

We will soon need to estimate how many residents we can accommodate in a small or medium sized container. For our calculations, we need to know the length l, the radius r, and the number of layers n of the container.

- The circumference of the kth layer (counting from top to bottom) would be:
  - $p = 2\pi(r - \frac{r}{n}k)$
- Thus, the total available surface would be:
  - $\sum_{k=1}^{n} 2\pi l(r - \frac{r}{n}k) = 2\pi l \sum_{k=1}^{n}(r - \frac{r}{n}k) = 2\pi lr \sum_{k=1}^{n}\left(1 - \frac{k}{n}\right) = 2\pi lr\left(n - \frac{1}{n}\frac{n(n+1)}{2}\right) = 2\pi lr\frac{n-1}{2} = \pi lr(n-1)$

Settlement Scoring Criteria

The following criteria would be used in the Principal Component Analysis to separate the generated settlements:

- Settlement capacity
- Allocated surface per resident (randomized at the start of every generated settlement)
- Percentage of the volume of the Settlement in which artificial gravity is present
- The minimum rectangular surface on which a piece extends. We will calculate this parameter 7 times, once for each piece. Obviously, we would want the surface to be as little (compact) as possible.
- The minimum rectangular surface on which humans perform their daily activities: living container, school/research area, hospital.
- The minimum rectangular surface which includes the other parts
- Total surface of the settlement.

These scores should be enough to determine the structure of a good settlement. Unfortunately, it is complex to differentiate settlements using all 13 scores, so the Principal Component Analysis should be extremely useful, letting us compare settlements using a smaller number of arguments.

Randomized parameters

At the start of each step (generating a new settlement), we need to randomize some parameters, in order to guarantee a different result every time.

- The number of people that the Settlement needs to host. Because of the modular nature of the Settlement, the number is between 200 and 1000 people.
- The space allocated for each human. It should be between $12$and $60\ m^2$.

Explaining the algorithm

We will be detailing each step in the process of making on settlement.

- Each piece we detailed earlier will have a number of connectable points (for example the 4-way connector will have four points that can be used to attach other components of the settlement).
- Thus, we can view the Settlement as a graph, with edges representing connections between parts, and nodes as the parts themselves, but with some properties (length and radius).
    - Of course, we have to make sure that there aren't two parts that would come in contact with each other.
- After receiving the other randomized parameters, the algorithm will calculate the number of pieces of each type it would like to use. We will detail this part below.
- Finally, the algorithm will begin by placing a 4-way connector in the origin of the XY plane. We will define a MIN-heap that does comparisons based on the Euclidian distance to the center of the settlement (the first 4-way connector placed). We will place the 4 connectors found on the first piece in the heap, in no particular order.
- Next, we will take the first connector out from our MIN-heap (the closest to the center), and find which pieces that we have to place can be mounted in that spot, without colliding with any existent ones.
- We will now actualize the number of pieces of any kind that still need to be placed.
- We also need to add to the queue the connectors of the new part that weren't yet attached to anything else.
- We will repeat these three steps until we will achieve a functional settlement, in relation with what we initially wanted.

Calculating the number of parts needed

- The 2-way living container has an available surface equal to:

$$S = \pi \cdot 30 \cdot 20 \cdot 4 \cong 7540m^2$$

The number of living containers needed is $\frac{no\_of\_humans \cdot surf\_per\_human}{S}$. We add one if the decimal part is different from 0.

- It is known that as little as $0{,}07ha$ of cultivated land can continuously feed one human. Since the agriculture container should have ten layers, each measuring $2m$ in height, each container can provide food for $\frac{\pi \cdot 30 \cdot 20 \cdot 9}{700} \cong 24{,}22\ humans$.
- Our solar panels should have an area of $600m^2$ each ($L = 30m, l = 20m$). Our solar panels will have a 20% efficiency. During standard tests, a $1m^2$ solar panel will output approximately 200W. Our settlement is placed in L1, almost always receiving direct sunlight. Even more, it is known that solar panels have a higher efficiency in colder environments. We have data from central Colorado, in which a $1m^2$ solar panel has its daily output close to $230\frac{W}{m^2}$. This is relatively great for Earth, so we will use it as a reference for our Space Settlement. In the same way, a $600m^2$ panel with a 20% efficiency should produce $240000\frac{kWh}{year}$. The yearly average consumption per human is $2674\frac{kWh}{person/year}$, so each panel should satisfy the needs of almost $\frac{240000}{2674} = 89{,}75 \cong 90\ humans$.

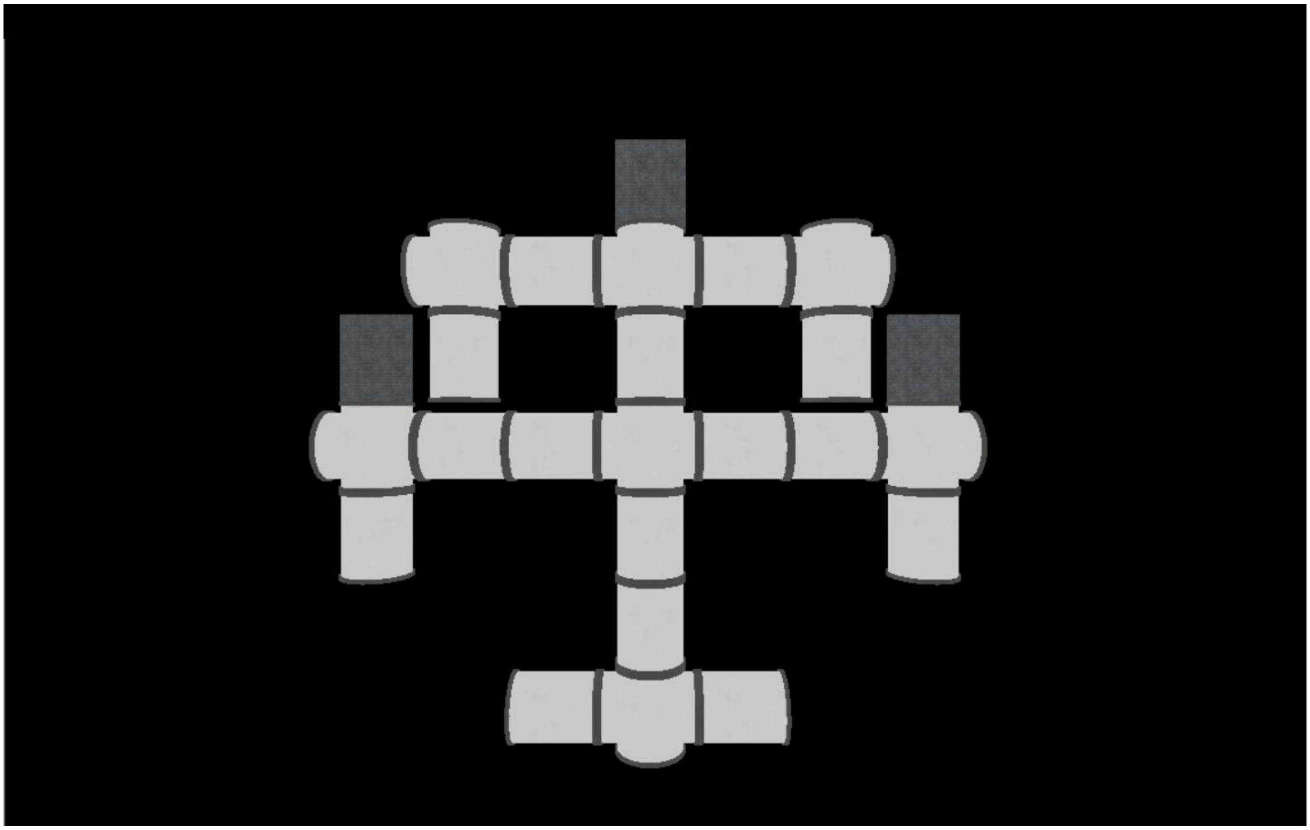
Example of a produced result by the generator

```
215 humans, 24 m sq per human
1 1 9 1 1 2 3
                                                7
                                            0 1 0 3 0
                                          7 4   5   6 7
                                          0 3 3 0 2 6 0
                                          3     3     3
                                                3
                                              3 0 3

215 24 0.325834 1 1 42 1 1 2 10 20 42 49
```

The graph on the right shows the top-down view of the settlement, in which every part has its number from the list previously shown. On the second line it is shown how many parts of each type (1-7) are used, taking into consideration that we have an infinite number of 4-way connectors (type 0). The very last line shows the 13 scores presented earlier, which will be used for our Principal Component Analysis.

We have rendered the above settlement:



Notice that all the 2-way containers (parts 1-6) aren't represented differently in the picture.

Observation: We can obtain interesting results by choosing a different data structure for holding the free connection points, rather than a MIN-heap. For example, a MAX-heap would create a degenerated graph (Settlement), since it always chooses the farthest free connection point in relation to the center. A normal queue would be equivalent with choosing a MIN-heap with a Manhattan distance-type comparison.

After enough experimentation, it became clear that we would want to use the Euclidean distance comparison instead of the less complex Manhattan Distance, for a reason we will explain below.

Principal Component Analysis of the 13 parameters

After running the generator ten times, as well as performing the Principal Component Analysis on the non-constant components, we have obtained the following Eigenvalue distribution:

Principal Component Analysis: ss capacity; spc per ... t; S other; S total

Eigenanalysis of the Correlation Matrix

Eigenvalue	6,0175	1,9850	0,9592	0,4565	0,3326	0,1377	0,0571	0,0406	0,0091	0,0047
Proportion	0,602	0,198	0,096	0,046	0,033	0,014	0,006	0,004	0,001	0,000
Cumulative	0,602	0,800	0,896	0,942	0,975	0,989	0,995	0,999	1,000	1,000

After running the generator for another 10 times, we have noticed that the difference in proportion is small enough to stop running the generator:

Principal Component Analysis: ss capacity; spc per ... t; S other; S total

Eigenanalysis of the Correlation Matrix

Eigenvalue	6,1822	2,1619	0,8799	0,3771	0,1605	0,1368	0,0596	0,0381	0,0041	0,0000
Proportion	0,618	0,216	0,088	0,038	0,016	0,014	0,006	0,004	0,000	0,000
Cumulative	0,618	0,834	0,922	0,960	0,976	0,990	0,996	1,000	1,000	1,000

↓	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14
		ss capacity	spc per resident	ratio space w/ gravity	S occp living	S occp school	S occp agriculture	S occp nucl	S occp hospital	S occp refinery	S occp panels	S occp important	S other	S total
1		723	42	0,200688	48	1	100	1	1	10	45	48	121	121
2		590	50	0,215469	36	1	72	1	1	3	35	36	81	81
3		858	16	0,129473	8	1	121	1	1	5	45	15	121	121
4		992	16	0,128407	28	1	132	1	1	12	50	28	132	132
5		596	13	0,175492	10	1	63	1	1	4	27	10	81	81
6		720	46	0,204303	27	1	100	1	1	8	36	54	100	121
7		857	28	0,161037	48	1	121	1	1	4	42	54	121	121
8		936	60	0,202731	45	1	132	1	1	4	45	45	144	144
9		265	43	0,327790	8	1	42	1	1	6	20	18	49	49
10		442	41	0,245460	8	1	56	1	1	4	36	28	81	81
11		458	58	0,260177	15	1	72	1	1	8	40	25	81	81
12		654	40	0,196131	20	1	90	1	1	2	36	28	99	99
13		778	37	0,172664	16	1	110	1	1	3	32	28	121	121
14		948	36	0,161121	24	1	132	1	1	10	44	24	132	132
15		653	16	0,164123	27	1	80	1	1	20	40	45	99	110
16		816	16	0,134753	8	1	110	1	1	6	45	21	110	121
17		976	48	0,183595	54	1	132	1	1	21	44	54	132	132
18		293	47	0,285334	5	1	42	1	1	8	18	18	49	49
19		443	14	0,197660	1	1	56	1	1	6	28	20	81	81
20		557	31	0,206980	20	1	72	1	1	2	50	35	90	90

From the cumulative proportions, we can safely say that we only need 4-5 components, out of the 10 non-constant ones.

Eigenvectors

Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
ss capacity	0,390	-0,109	-0,120	0,031	-0,007	-0,303	-0,268	0,265
spc per resident	-0,042	0,633	-0,041	0,528	0,067	0,043	-0,510	0,124
ratio space w/ gravity	-0,316	0,355	0,204	0,315	0,090	-0,260	0,623	0,001
S occp living	0,267	0,424	-0,009	-0,523	0,624	-0,187	-0,013	-0,061
S occp agriculture	0,390	-0,043	-0,071	0,158	-0,124	-0,377	0,302	0,587
S occp refinery	0,162	-0,107	0,959	0,024	0,034	-0,067	-0,165	0,008
S occp panels	0,377	-0,044	0,028	0,239	0,280	0,756	0,231	0,211
S occp important	0,216	0,515	0,088	-0,399	-0,619	0,240	0,170	0,039
S other	0,392	-0,009	-0,090	0,237	0,153	-0,095	0,265	-0,534
S total	0,392	0,028	-0,026	0,221	-0,311	-0,136	-0,089	-0,485

- By analyzing the PC1 eigenvector we can conclude that the *Settlement Capacity*, *the surface occupied by the solar panels*, *the surface occupied by the Agriculture containers*, *the surface occupied by some components (Agriculture, Nuclear energy, Refinery, and Solar Panels – "S other")*, and the *total surface occupied by the Settlement* are strongly related.
- Then, we can compare settlements based only on the *Settlement Capacity*, *Surface per Resident*, *Ratio of volume with gravity in relation with the whole settlement volume*, *surface occupied by the living containers*, and the *surface occupied by the daily purposed containers*.

Based on the previous deductions, we think that the 13th generated settlement is a great contender for a good Space Settlement design:

```
778 humans, 37 m sq per human
4 1 33 1 1 2 9

          6   1
    0 0 3 0 5 0 1
    4 7   6   1
0 7 0 7 2 0 3 0 3 3 1
7   3     7   3 3
7 7 0 3 3 0 7 7 0 3 3
3   3 3   3   3 3
0 3 0 3   3   3 3
  0 3 0 3 0 3 0 3
      3   3   3 0
      3 3 0 3 3

|
778 37 0,172664 16 1 110 1 1 3 32 28 121 121
```

Artificial gravity

Elysium must have an artificial gravity system in order to simulate Earth conditions. This can be done by spinning the cylinders around their axis. As we know, the normal acceleration is equal to $a = \frac{v^2}{r}$, where v is the speed of the circular rotation, and r is the radius of the cylinder. On earth, the gravitational acceleration is approximately $9.81\ \frac{m}{s^2}$, thus, this will be the emulated artificial gravity. We will emulate the artificial gravity in the living areas cylinder, with a radius of approximately 20m. Therefore, the speed will be $v = \sqrt{a * r} = \left(\sqrt{9.81 * 20}\right)\frac{m^2}{s^2} \approx 14.007\ \frac{m}{s}$.

Procuring the needed materials and costs

The needed materials for the settlement will be transported with an Atlas V rocket from Earth. Because of the ship's modularity, each component will be assembled in space with less effort. We could also take advantage of the fact that Mars' dust could be compressed into bricks, which could be helpful if we decided to build a Mars settlement or to enlarge Elysium. Here is the cost table for a living container (which is a cylinder with a height of 30m and a radius of 20m):

Aluminium507,000 kg (5cm thickness)	912,600 USD
Lead 1,282,000 kg (3 cm thickness)	2,649,054 USD
Polyurethane (used for thermal shielding) 200,000 kg (5 cm thickness)	394,236 USD
10 Server Computers	50,000 USD
Solar Panels (1 panel per 90 humans) 600 m^2, 20% efficiency	231,990 USD
Total	4,237,880 USD

IV. Life Support

Atmosphere

Requirements

- the partial pressures of the gases found in atmosphere must be in certain ranges, in order to assure good respiratory conditions, as well as proper conditions for plants growth
- partial pressure of O_2 needs to be above the required minimum for the alveoli of the lungs, but below a certain level to prevent hyperoxia and also an increase of the number of opportunistic bacteria that use oxygen; the value of p_{O_2} mentioned in the table is slightly higher than the value p_{O_2} from Earth (22,7kPa)
- N_2, an inert gas which is prevalent in Earth's atmosphere is a safety measure in case of accidental pressure drops; on Earth, nitrogen is part of a cycle which leads to a constant interchange between the atmosphere and living organisms; to partially reconstruct this cycle would be benefic for our settlement because it would keep the nitrogen level mostly constant, as well as assuring an efficient agriculture

Reconstruction of nitrogen cycle:

- we will use Rhizobium and Cyanophyta as nitrogen-fixing bacteria that contain the nitrogenase enzyme that catalyze the reduction of N_2 to NH_4^+cation
- nitrifying bacteria (Nitrosomonas) will help the oxidation of NH_4^+toNO_3^-
- part of nitrates will be assimilated by plants (the species we use are mentioned at Agriculture section), while the rest of them will be turned to atmospheric nitrogen by denitrifying bacteria (thiobacillusdenitrificans).

- carbon dioxide can lead to irreparable effects when its partial pressure exceeds $1kPa$, so it is safe to maintain a low level of $0,3kPa$, which is necessary for the plants' growth

p_{O_2}	25 kPa
p_{N_2}	27 kPa
p_{CO_2}	0,3 kPa
p_{H_2O}	1,0 kPa

- the total pressure is 53,3kPa, which is nearly half of the Earth's atmosphere

Processes

Even though there are pressurized oxygen storage tanks stored in the settlement, we will also use other methods of producing O_2:

- **Electrolysis of water**
  - we will use it as the main method; the reactions taking place are:

$$\text{Cathode: } 2H2O(l) + 2e- \rightarrow H2(g) + 2OH-(aq)$$
$$\text{Anode:} 2OH-(aq) \rightarrow \frac{1}{2}O2(g) + H2O(l) + 2e-$$

- **Decomposition of $KClO_3$**
  - advantageous method because of its cheap materials; when potassium chlorate is heated at approximately 150 °C with small quantities of MnO_2, which is a catalyst, the following reaction occurs:

$$KClO3(s) \rightarrow KCl(s) + \frac{3}{2}O2(g)$$

(used as a back-up method in case of major incidents)

- **O_2 and CO_2 plant cycle**
  - CO_2 is absorbed by plants, and by photosynthesis process, O_2 is produced:

$$6CO_2 + 6H_2O \xrightarrow{hv} C_6H_{12}O_6 + 6O_2$$

  - The Sabatier reaction is very helpful in the given conditions, using only H_2 that was formed in water electrolysis (H_2 is quite reactive, so to prevent any incidents it is preferred to get rid of it) and CO2, which level we want to reduce as well

$$CO2 + 4H2 \xrightarrow{400\ °C} CH4 + 2H2O$$

Gas Monitoring System

- in order to maintain the safety of the settlement, it is necessary to have a gas detector that can measure the level of certain toxic gases, as well as an oxygen deficiency gas monitor:
  - H_2S monitor
  - O_2 monitor
  - CO and CO_2 monitor
  - monitor for low levels of combustible gases

Temperature

- There are two aspects which we are concerned about: the ambient temperature and the spacecraft's component system temperature.
- The thermal control system has two roles:
  - to protect the equipment from overheating, by thermal insulation from external heat or by heat removal from internal sources
  - to protect the equipment from too low temperatures

Water Purification System

- Total water recycling is absolutely necessary, because the glacier from the South Pole cannot provide enough water to cover the water needs. Also, water recycling is much easier to perform, as well as implying a smaller number of risks than collecting water from Mars.
- Water from any source(the one produced by the Sabatier reaction, urine, moisture excess collected from air, garbage, leaves etc.) will be collected and then purified
- Purification consists of physical, chemical and biological processes:
  - the first step is to separate organic solids: artificial gravity is used for an efficient decantation
    - these organic residues are then decomposed by bacteria
  - then, nitrates and phosphates need to be taken out of the water – this can be done by taking the water, at a very high pressure, through a semi permeable membrane
  - the last step consists of methods against microbiological contamination: ultraviolet rays and heat are used to prevent the presence of any microorganisms(water is cooled and then heated, a big temperature gap being a good method to destroy the bacteria)

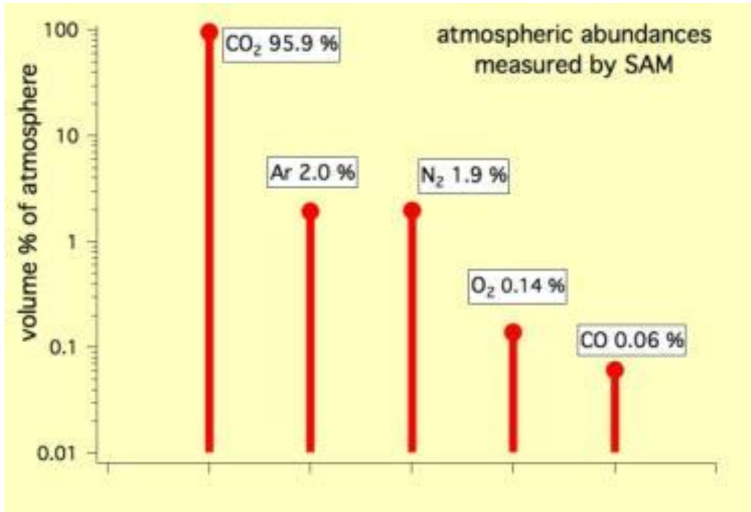
V.In situ resource utilization

The aim of this chapter is to discuss about the gases and minerals found on Mars and the methods we use to exploit them in order to sustain life on settlement.

Gas analysis

Prevalence of CO_2

The main component of Mars atmosphere is CO_2, with a molar percentage of 95.9%, as measured by Sample Analysis of Mars Suite Investigation. Despite of being fatal for human body exposed for 5 minutes at concentrations as low as 150 g/m^3, we still can find some useful applications.



[source: https://photojournal.jpl.nasa.gov/catalog/PIA16460]

While in its hemisphere's winter, each pole is in total darkness and the temperatures drop very low. As a consequence, 25% of CO_2 condenses into solid CO_2, called dry ice. When the exposure to sunlight starts, the solid CO_2 sublimes back. This process causes an important annual variation of CO_2 concentration and atmospheric pressure.

Dry ice is an efficient cooling agent. It will be collected and used on the station as a very cheap method to preserve food.

We dispose of two processes for CO_2 conversion:

1. Sabatier reaction $CO2 + 4H2 \xrightarrow{400°C} CH4 + 2H2O$

2. CO_2 splitting $CO2 \xrightarrow{Zr\ oxide} CO + \frac{1}{2}O2$

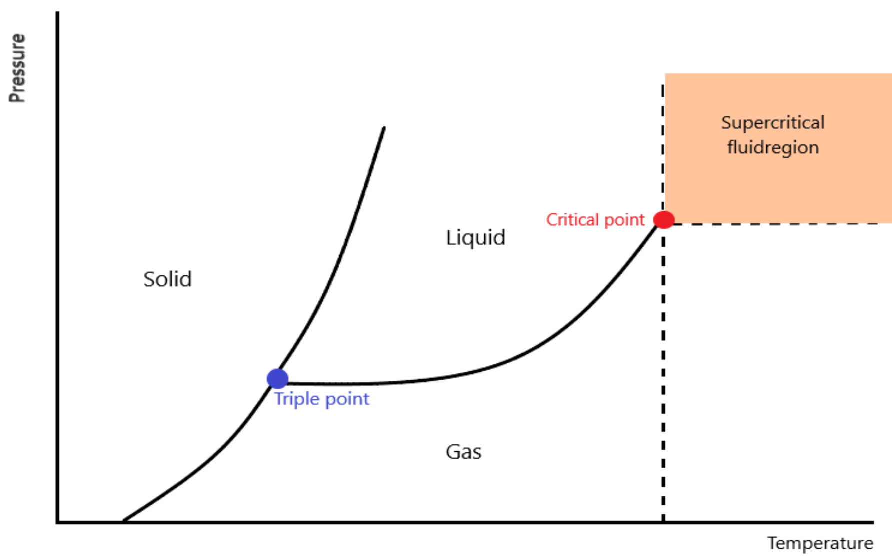
Ore resources

Mars provides useful ore deposits, including many elements that are expensive on Earth. Copper, chromium, iron and nickel are found in various combinations, forming minerals. Elements that do not bind easily with others, incompatible elements, can also be found there: niobium, lanthanum, neodymium and europium.

Extraction

To extract these useful elements, we need a very strong solvent and for that we make use of the carbon dioxide found in atmosphere.

A supercritical fluid is any substance at a temperature and pressure above its critical point. It can effuse through solids like a gas and dissolve materials like a liquid.



- CO_2 is the most common supercritical solvent. For this, we need to compress it at a pressure of 73 atmospheres and to heat it to 35°C.

- One of the elements that can be dissolved easily by CO_2in its supercritical stateis magnesium, which was discovered on Mars rocks. Combustion of Mg is a method to produce rocket fuel, so this solvent could have many useful applications.

- Supercritical CO_2 can also be used to produce water. Certain rocks contain important amounts of hydrogen. When hydrogen goes through this solvent, a chemical reaction takes place and H_2O is produced.

VI.Radiation Shielding

Radiation, defined as a transmission of energy in the forms of waves or particles, is one of the biggest threats for the settlement's inhabitants, so a careful analysis has to be done.

Space radiation is classified in three types:

- particles from solar proton events (SPE)
- galactic cosmic rays (GCR)
- particles that kept in a Van Allen radiation belt (a region where energetic charged particles are trapped in a Earth's magnetic field)

Since the time spent passing the Van Allen belt is very short, it won't produce any significant effects (the radiation is comparable or even less with the radiation a human is exposed to during a computed tomography scan).The main concern includes the first two types of space radiation, which we will describe in depth.

- Cosmic rays are the dominant source of radiation. They are composed of high-energy protons and atomic nuclei which have their origins outside our solar system or from the sun.

Radiation Limits

Roentgen is the unit for the quantity of radiation.The sievert (Sv) is a derived unit that refers to the ionizing radiation dose, also being the most frequently used way to measure health effects on human body when exposed to low levels of ionizing radiation.

Category	Radiation limit
Specialized workers	50mSv
General Public	1mSv
Minors and pregnant women	0.5msV

The dose that can kill a human within 30 days with a risk of 50% is approximated to be 4-5sV.

To avoid any medical complications, the settlement needs to respect the limit of 1msV set for the general public.

Materials used

Considering that galactic cosmic rays cause atoms they pass through to ionize, they can theoretically pass through a spacecraft and through the skin of human as well. Choosing a proper material for the radiation is essential in order to keep the inhabitants away from any danger.

For example, lead is one of the best materials to use against electromagnetic radiation (gamma and X-rays), because of its high atomic number, which means it has a high number of electrons. We know that the electrons block the many gamma and X-ray particles that try to pass through a lead barrier. The effect can be multiplied by stacking multiple lead barriers.

Alpha and beta radiation can be easily blocked by a plastic shield. Lead is ineffective, since this type of rays produces unwanted secondary radiation when passing through elements with a high electron number.

In order to block neutron radiation, again lead is undesirable on its own, because the neutrons are uncharged, and thus pass through dense materials. Low atomic number elements or compounds are preferred for shielding against this type of radiation, including hydrogen and water, because they will have a higher probability of interacting with neutrons. Unfortunately, when interacting with the neutrons, hydrogen produces secondary gamma rays, which have to be blocked by a lead barrier.

VII.Gas Geysers

Martian Geysers are sites of gas and dust eruptions that become active during the spring thaw; they are found in the south region. Next, considering that on long-term we plan to expand our settlement, we want to approximate how many people the geysers can support.

It is known the molar distribution of the gas mixture produced by the geysers:
95.32% CO_2, 2.7% N_2, 1.6%Ar, 0.13% O_2, 0.08% CO.

We know that the temperature near the South Pole of Mars is 140K, and the pressure varies between 6.7 and 8.8 mbar.

With the weighted average formula, we calculate the average molar gas of the mixture:

$$\bar{\mu} = \frac{\sum \mu_i \vartheta_i}{\vartheta_i} = \frac{44 \cdot 95.32 + 28 \cdot 2.7 + 40 \cdot 1.6 + 32 \cdot 0.13 + 28 \cdot 0.08}{100} = 43.2 \frac{g}{mol}$$

The root mean squared speed of a molecule is calculatedas $\sqrt{\frac{3RT}{\mu}}$, from which we can obtain the average speed of a molecule from the gas to be:

$$v = \sqrt{\frac{3 \cdot 8.31 \cdot 140}{43{,}2 \cdot 10^{-3}}} = 284{,}23 \frac{m}{s}$$

From some satellite images we can infer that the geysers can be found over many hundreds of kilometers, with at least one geyser for each 1-2 km. It is also stated that their diameter is between 15 and 40 m. For proper approximations, we choose the average diameter as 20m.

Then, the volumetric flow rate is $Q = \pi r^2 v = 357173{,}9 \frac{m^3}{s}$

Using the Clayperon-Mendeleev formula for gases, we find the molar volume in comparison to the one calculated on Earth:

$$pV = \vartheta RT.$$

For $T = 298K$ and$p = 101325Pa$,

$$\frac{V}{\vartheta} = \frac{RT}{P} = \frac{8{,}31 \cdot 298}{101325} = 24{,}4 \cdot 10^{-3} \frac{m^3}{mol}$$

Near Mars' southern pole, the molar volume is:

$$V_{mol} = \frac{V}{\vartheta} = \frac{RT}{P} = \frac{8{,}31 * 140}{875} = 1{,}31 \frac{m^3}{mol}$$

Then, the molar debit is $D = \frac{Q}{V_{mol}} = \frac{357173{,}9}{1{,}31} = 272651{,}8 \frac{mol}{s}$

Finally, we can calculate the molar debit of CO_2 and N_2, which we will note with $D_{CO_2}$and D_{N_2}.

$$D_{CO_2} = D \cdot p_{CO_2} = 272651{,}8 \cdot 0{,}9532 \cong 259{,}89 \frac{kmol}{s}$$

$$D_{N_2} = D \cdot p_{N_2} = 272651{,}8 \cdot 0{,}027 \cong 7{,}3616 \frac{kmol}{s}$$

Presuming that a human breaths 18 times per minute and the inhaled volume of air is of 1.5l, it would be need

$$V_h = 24 \cdot 60 \cdot 18 \cdot 1{,}5 = 38{,}8 \frac{m^3}{day/human}$$

Using the previous calculations, it can be approximated that the following volume of N_2 will exit the geysers per day (which means 86400 seconds)

$$V_t = D_{N_2} \cdot V_{mol} \cdot 86400 = 833215334m^3$$

Using the methods described earlier for obtaining oxygen and considering that air composition is approximately 80% N_2 and 20%O_2, we could probably produce enough air for

$$\frac{\frac{100}{80}V_t}{V_h} = \frac{1041519168}{38{,}8} = 26\ 843\ 277\text{humans}$$

VIII.Risk event analysis for water collection

Even if replenishing our resources from Mars is a risky process, it is an absolutely necessary trial. A lander will be tasked with collecting the water from the glaciers on Mars' South Pole.The following events have been included for the risk tree:

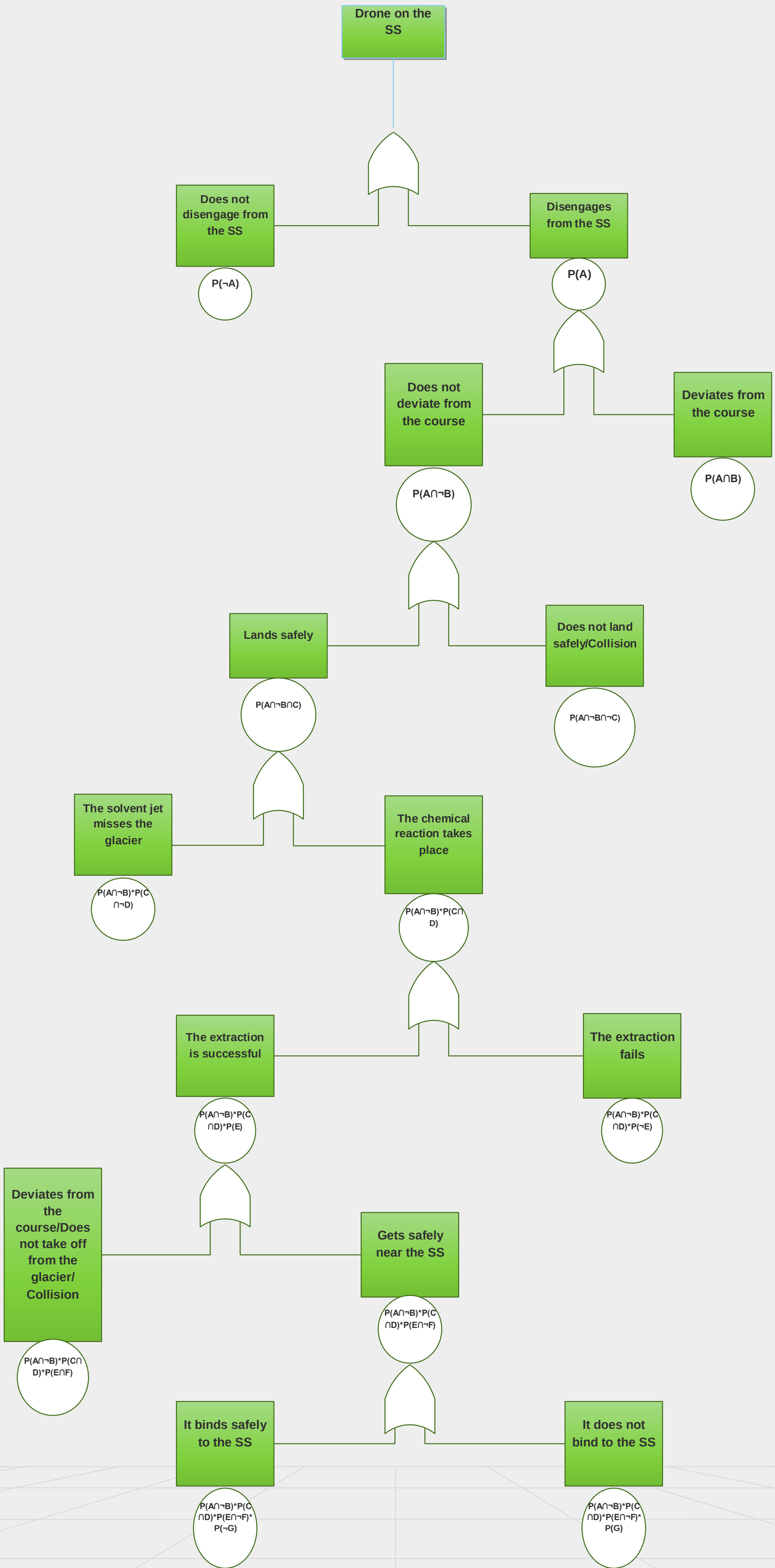
- (A) The lander's separation from the settlement
- (B) Following the trajectory toward the glaciers. Not following the way may lead to us losing the lander, which would greatly hinder our capability of gathering enough water for various activities.
- (C) The landing, taking into account the functionality of the jets, which are supposed to slow the lander down, since Mars' atmosphere isn't thick enough to permit a parachute landing. Also, a malfunctioning sensor can lead to the jet gaining too much or too little thrust, deviating it from its course.
- (D) Spraying CO_2 in its supercritical state over the glacier
- (E) Collecting and storing the water
- (F) Returning on the same path to the Space Settlement. Again, if the jets do not function properly then the mission may be interrupted
- (G) Safely attaching/binding the lander back on the settlement.

Observation: Since failing almost any event, with the exception of the first (lander not detaching) is considered an abrupt mission ending, the risk event analysis tree will closely resemble a degenerated tree.

We will denote P(A) as the probability event A will take place and with P(¬A) the probability that event A will not take place. This means that P(¬A) denotes the probability that an event complementary to A will happen. Thus we have P(A)+P(¬A)=1.

LetP(A ∩ B) be the probability that both event A and event B will happen, with A and B being independent events. We know that $P_B(A) = \frac{P(A \cap \mathrm{B})}{P(B)}$ where $P_B(A)$is the probability that event A will happen with the condition that event B will also happen. As A and B are independent events$P_B(A) = P(A)$. Thus we obtain $P(A \cap \mathrm{B}) = \mathrm{P(A)P(B)}$. All events from our diagram are independent, so we are free to use the above formula.

Fault Tree Analysis Diagram



IX. Biology

- *organisms that live in a rich CO_2 atmosphere on Mars that may present a threat to humans:*

Based on the recent screenings (June 2018) made on Mars (ExoMars Trace Gas Orbiter), the Martian atmosphere is likely to have methanogenic bacteria (eg: Methanococcus, Methanobacterium, Methanospirillum, Methanosarcina) because of the methane concentration in the Martian atmosphere that is fluctuating seasonally. The presence of these chemosintetizing species could be the cause of these variations.

There may be other species of anaerobic microorganisms (such as SpinoloricusCinzia species). Extremophile species resist very high or low temperatures, low oxygen or high $CO2$ concentrations, and can, under unfavorable environmental conditions, generate spores through the sporogenesisprocess.

If we assume that there is water under the rocks from Mars, then the possibility of living on Mars becomes certain, and the organisms that may exist in the presence of rocky water would probably be similar to those found in the caves on Earth: viruses, bacteria, protozoa, or fungi. Viruses, bacteria, and some molds are pathogens that are harmful to humans either by direct contact (blood, saliva, etc.) or by inhalation – in a few cases.

Generally, methanogenic bacteria (found in stomachs of ruminants and which break down the cellulosic wall of plant cells) do not have a direct pathogenic potential for humans, but can cause disease by destroying good microorganisms that naturally populate our intestines. Parasite protozoa can also cause many diseases.

- *anaerobic organisms which arrive with humans on Mars and subsequently adapt to the atmosphere:*

The species of extrinsic or chemosynthetic bacteria previously referred can reach Mars with humans and adapt to the atmosphere of Mars. Several experiments, regardingMethanogenic bacteria, in the absence of oxygen, consume the CO_2 atmosphere and generate water and methane ($4H_2 + CO_2 \rightarrow CH_4 + 2H_2O$). It would be extremely useful, for that they can produce water.

- *diseases that humans are prone to developing in a 0g environment:*

As other scientists have concluded, the complete lack of gravitation or less gravitational attraction over a longer period of time causes cardiovascular diseases because the endothelium (the tissue that lines the interior surface of the blood vessels) ages faster.

It was determined that in cells living in an atmosphere with a low gravitational pull had a high level of cytokines – molecules associated with inflammation. This process leads to the amplification of oxidative stress, resulting in inflammation in the endothelial cells, leading to atherosclerosis and cell senescence (biological aging).

Furthermore, there are over 1000 genes that activate under oxidative conditions and cause cellular aging. Consequently, other diseases associated with aging, such as neurodegenerative diseases like Alzheimer's or Parkinson's disease, may occur.

In the absence of natural sunlight, we can assume that people on the settlement will develop eye diseases such as cataracts. It is completely possible in the long run the pupil may enlarge to allow a larger amount of light to be projected on the retina. Thus, in a few generations all humans on the settlement could have black eyes. Lack of natural light would have an impact on the assimilation of vitamin D3, which will cause serious calcium fixation problems in bone, and hence decalcification-specific diseases such as osteoporosis, rickets or bone fragility. Also, the excretory system would be drastically impaired, since the sensation of urination occurs when the bladder is 1/3 full. A low gravitational pull implies people should urinate without knowing when their bladder is full. This means that the possibility of urinary system diseases(glomerulonephritis – kidney blood filter diseases for example) occurring drastically increases.

X. Life on the Settlement: Social structure

On the settlement ship, there will be living between 200 and 1000 people, depending on the random parameters. 80% of the people will just be civilians, and 20% will be the scientific and maintenance crew. Out of the 20% of the people, 5% will be scientists responsible for conducting experiments in the microbiology laboratory and developing theories in the study room. 2.5% people will be doctors and 5% of people will be nurses. 3% of the people will be tasked with keeping the settlement clean. 2% people will be mechanics that will keep the ship functional. 1.5% will be IT technicians will be monitoring the ship. 1% of the people will represent the command centre of the ship.

A healthy life for all the inhabitants of Elysium is one of the priorities of our mission. Thus, doctors from a wide variety of medical fields must be recruited, such as psychologists, nutritionists, dentists and surgeons. Members of the ship will also have a mandatory monthly health inspection, in order to prevent the risk of dangerous diseases, as well as contain any possible contamination.

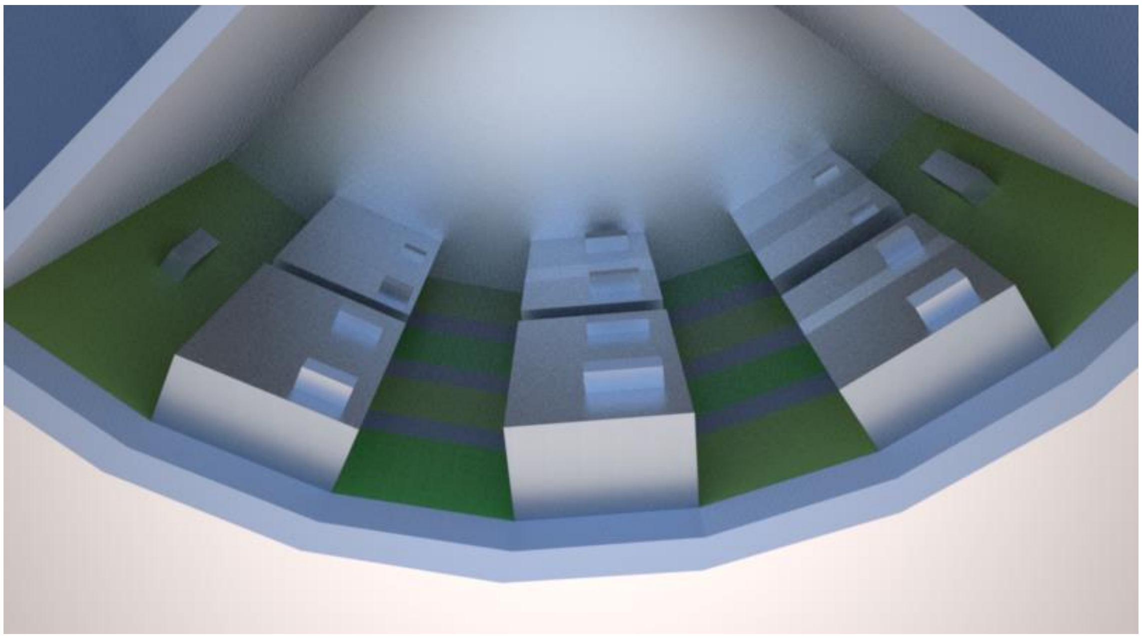
The entire population will be carefully selected on Earth. Special training courses on our planet will be held, of lengths of over 4 years in order to be able to live an optimal life on the ship. The staff will also be required to take additional courses in order to adapt their skills to the difficulties of space.

Room management

School and research area

In order to provide the new generations of settlement inhabitants with a proper education, as well as providing the science team with a dedicated space for research and theory development, the decision to include such a space seems necessary. For this type of room, a small 2-way living container, with a length of 30m and a radius of 20m is used.

The cylinder will be divided in 4 sections, each being identical to the other. Each section will contain 3 rows with 2 buildings each, which will house the spaces for studying. Every building will be a parallelepiped with a length of 10m, a width of 5m and a height of 4m, housing two classrooms of 16 students each. Each section will also have a campus which will be used for relaxation activities, as well as sports. Here is an image of a section:

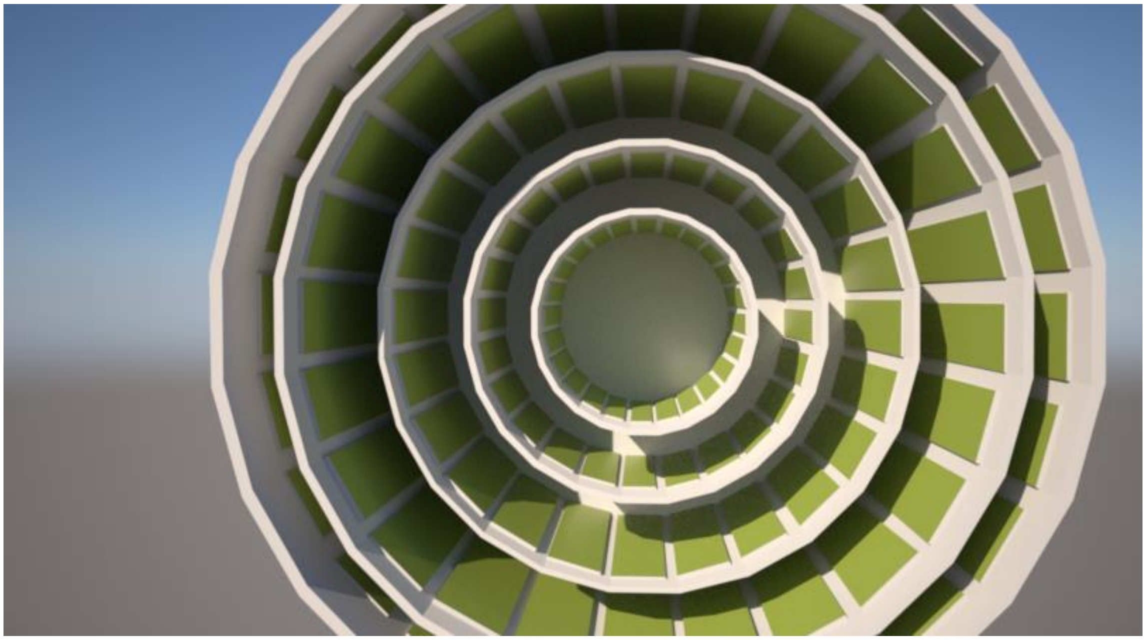


Agricultural area

The agriculture of the space settlement is bound to be very productive in order to provide food for the entire population. This is why the best decision is to mount several concentric cylinders in order to provide a larger agricultural area.

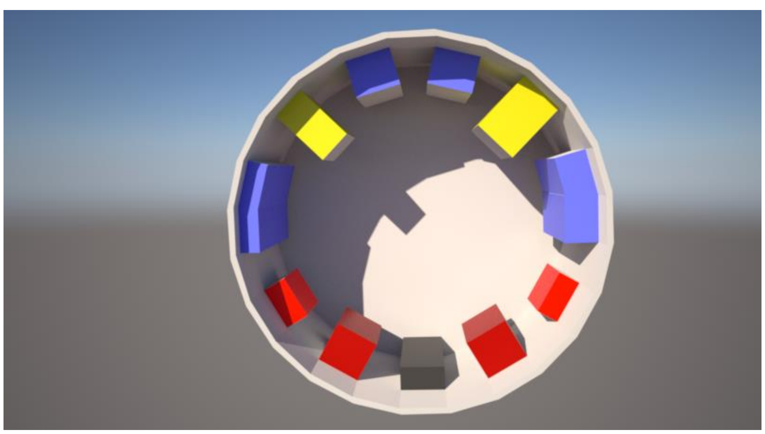
We have also decided that a hydroponic agricultural system would be fitting for the purpose of growing food in space, as it would also provide a better ratio of food produced per square meter. Some tests have shown that production compared to normal methods could be 11 times more efficient. We will use a nutrient film technique, meaning that a continuous flow of nutrients and water will run over the plants' roots.

Here is an image of an agricultural area cylinder:



Nuclear energy area

Each chamber will contain one nuclear generator capable of having a power of 400 MW. These will not be covering the population’s needs, instead, it will cover the energy needs of the Elysium itself.

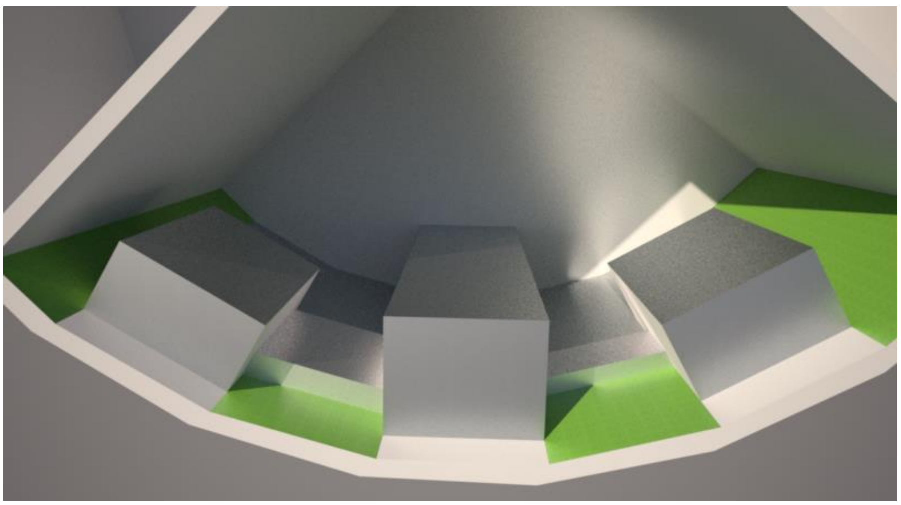


The black cube represents the control room. The red zones represent the main fission reactors, the blue zones are containers with cold water, and the yellow zones are handling the water evaporation. Because the water vapour is not radioactive, we could recycle the water and use it in the cooling of the reactor.

Hospital area

The hospital area will be responsible of monthly medical controls for the inhabitants of Elysium. These controls will be required in order to prevent unseen injuries or diseases, as well as prevent widespread disease.

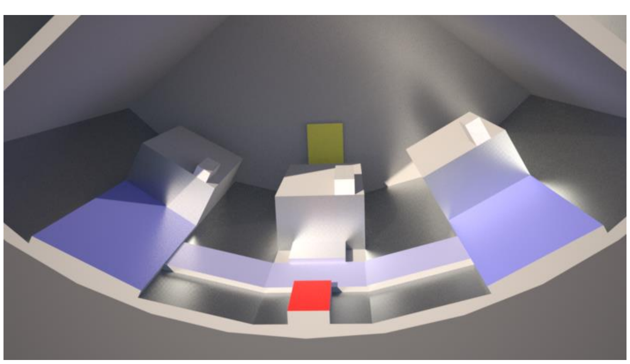
The cylinder containing the hospital will be divided in 4 sections, in order to avoid excessive waiting lines, as well as keep the hospital as organized as possible. Each section will contain three main buildings, as well as green space for relaxation.



Air and water refinery area

In this area, water and air will be recycled, because we have limited resources in space.

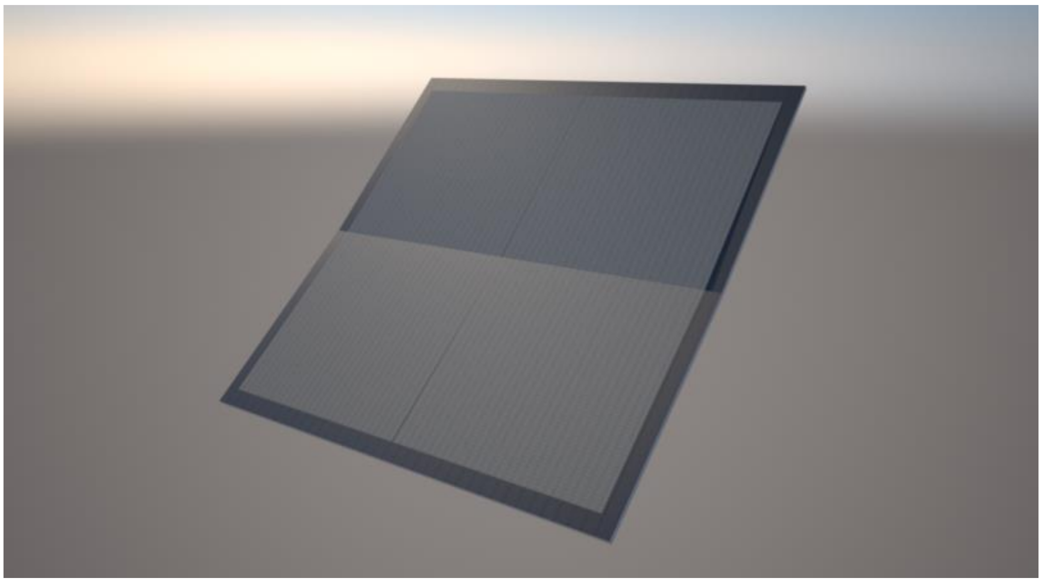
The blue zones represent the water tanks. Water will be circulated to the refining buildings. Air will be circulated through the yellow ventilation system through the back of the buildings. Purified air and water will be circulated through the red zone.



Solar panels

Solar panels will be responsible for providing energy for the ship, in conjunction with the nuclear energy task.

Every living cylinder will be provided with a solar panel with an area 600 m^2,



Letter to my friend John from Earth

Dear John,

How are you? I was glad that you could respond to my last email! How is school? I have just finished 11th grade here on Elysium and I'm entering my final year! I hope you'll send me some pictures of the flying car you've just bought, would love to see how those things look!

Yesterday was an extraordinary day, I had my first trip to Mars. It was so exciting and I took a lot of pictures! I also had a little contest with my friends on Elysium on who would build the biggest dust castle on Mars! I won! We also visited the Mars' Pioneers museum, where we saw the first ship to land on Mars and also an old astronaut suit used by the first man on Mars!

When I was back on Elysium, I was also excited to find out that a new sports area would open on one of the research parks, with basketball and soccer courts! The same day I played a game of basketball with my friends! I would really like to play a game with you too!

Well, I am really glad to respond to your email! Maybe one day, you'll come on Elysium for a visit?

Your friend,

Eli

Further reading

The C++ code used for the settlement generator can be found here.

The eigenanalysis code can be found here.

The generated settlements can be found here.

XI.References

Wolfram Mathematica, Minitab, Valve Hammer, Google Sketch-Up

"Numerical calculus elements", Demidovitch

http://www.math.tamu.edu/~dallen/linear_algebra/chpt6.pdf

„The coefficients of the characteristic polynomial in terms of the eigenvalues and the elements of an n × n matrix", Bernard T. Brooks

https://en.wikipedia.org/wiki/Atmosphere_of_Mars

https://www.space.com/16903-mars-atmosphere-climate-weather.html

https://www.lpi.usra.edu/publications/reports/CB-1089/taylor.pdf

https://www.spacefoundation.org/news/space-foundation-applauds-successful-placement-research-spacecraft-mars-orbit

https://www.space.com/30302-lagrange-points.html