What is the formula for the sum of squares?

With the support of many euristic arguments.

I assume that the question should be: What is the the sum of the squares of the natural numbers from 1 to n.

The answer is

$$(1) \qquad S(n) = \sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6} = \frac{1}{6}\,n + \frac{1}{2}n^2 + \frac{1}{3}n^3$$

Since the antiquity, the formula is known and has been derived in various ways. Probably, only the theorem of Pythagoras has more demonstrations: I may quote, among the formulas I know, the "simplest algebraic formula", Faulhaber's formula and derivatives, recurrence formulae, a number of finite differences and derived formulae etc. – up to the "*mathematical induction principle*".

The last quoted method shows that if one sums the squares up to (n+1), that is, adds (n+1)^2 to formula (1), he can rearrange the terms in such a way as to reproduce the (1) with (n+1) now replacing n everywhere. In other words, if the formula is valid for n, it is valid for n+1. Then one shows that the formula is valid, for example, for n=2 (and it is, because 1/3+2+8/3=5 =1+2^2 = S(2)), which entails that it is valid for n=3 and then n=4 etc. to n=infinity. I have sketched the method because I found it irritating. Nobody tells us where formula (1) comes from. Thus, the exercise I have outlined is an excellent example to illustrate the mechanisms of the principle of mathematical induction, but not to produce formula (1).

I will now give the most immediate algebraic method to find the complete formula, and the simplest graphical way, understandable to an intelligent kid, to get an intuitive grasp of formula (1), for the case n=4. Then, if one wants, one can apply the mathematical induction, but - frankly – I think that euristically one can accept the formula without further ado.

1. An exercise in elementary algebra.

Algebraically, **the first thing to do is to realize/demonstrate that the formula for S(n) is a polynomial of the third degree in n.** If we know **(1),** we just look at it, and there we are. But what if we don't? Well, the first thing to do is to trust our luck and bet on the polynomial form, which is the simplest one. Then, if it will not work, we will try something more complicated.

There is something interesting about the powers of natural numbers. Let's look at the differences between consecutive terms (the so-called first differences, for us Δ(1)), then the differences between the consecutive first differences (the so-called second differences, for us Δ(2)) and so on.

For example, suppose we have the power 1

n	1		2		3		4		5
Δ(1)		1		1		1		1	
Δ(2)			0		0		0		

Clearly, it is useless to go any further. From the second difference onwards all differences Δ(i) will be zero.

Let's try with power 2 (multiplied by any coefficient, such as **a**)

n^2	0		1a		4a		9a		16a		25a
Δ(1)		1a		3a		5a		7a		9a	
Δ(2)			2a		2a		2a		2a		
Δ(3)				0		0		0			

As we see, the coefficient **a** does not change the fact that the third differences (and all higher differences) are zero.

Let's try with power 3 (multiplied by any coefficient)

n^3	0		1b		8b		27b		64b		125b
Δ(1)		1b		7b		19b		37b		61b	
Δ(2)			6b		12b		18b		24b		
Δ(3)				6b		6b		6b			
Δ(4)					0		0				

Here we see that again the coefficient b does not change the fact that all differences higher that the third ones will be zero.

As an aside I note here that a smart pupil (with some rudiments of calculus) could tell that the constant 2, which appears in the differences of the squares, and the constant 6, which appears in the differences of the third powers, and that zero after the second differences for the squares and *after the third differences for the cubes, all sound familiar. In fact, they are the same coefficients which appear in the derivatives of x^n.*

Let's now explore what kind of polynomial is one which expresses the sums of the squares. We calculate the first few results by direct sum, 0+1+4+9+16 + 25…., which give: S(1)=0, S(1) = 1; S(2)= 5; S(3) = 14; S(4)=30; S(5) = 55….

S(n)	0		1		5		14		30		55
Δ(1)		1		4		9		16		25	
Δ(2)			3		5		7		9		
Δ(3)				2		2		2			
Δ(4)					0		0				

The coefficients make the 6 disappear, but we can deduce that the highest power in our polynomial is n^3. In fact, the lower powers will have disappeared earlier.

Again, a smart pupil, might notice that if we were to do the integral of x^2 from 0 to n, the answer would be (x^3)/3, which would already give the dominant term of our polynomial (1). For very large n, the quadratic and linear terms would pale into insignificance. For example, the three terms for n=120 are: n/6 = 20; (n^2)/2 = 7200; (n^3)/3 =576000 (80 times the second power contribution). The integral, in fact, is the limit of a sum, as the present cases shows.

Our polynomial will be S(n) = a + x n+ y n^2 +z n^3, and we will immediately dispose of a, because S(0) = 0 and therefore a=0 and

S(n) = x n+ y n^2 +z n^3.

Here, x,y,z are the three unknown coefficients of n, n^2, n^3. No more coefficients are needed. Since x, y, z will be the constants for our future formula, we just write three equations for n=1, n=2, n=3 and, recalling that S(1) =1, S(2) = 5, S(3)= 14, we will just solve the algebraic system:

$$\begin{cases} x + y + z = 1 \\ 2x + 4y + 8z = 5 \\ 3x + 9y + 27z = 14 \end{cases}$$

We have now is a simple linear algebraic system in the unknowns x,y,z. One can solve it in whatever way he/she prefers, for example by using Cramer's rule (1750), my favorite for theoretical purposes, or Gauss method (which works better for practical purposes) or whatever. I hope I will be forgiven if I will not show the straightforward calculations. *(Incidentally, Cramer must be pronounced in the French way, as Gabriel Cramer was born in Geneva, a French Swiss).*

What matters is that whatever method you use, you get **x=1/6, y=1/2, z=1/3**, and therefore the result, lo and behold, is:

$$(1) \qquad S(n) = \sum_{i=1}^{n} i^2 = \frac{1}{6}\,n + \frac{1}{2}n^2 + \frac{1}{3}n^3 = \frac{n(n+1)(2n+1)}{6}$$

A question is why should one reduce a simple three terms polynomial into the strange formula on the extreme right. The fact is that the last formula is more telling in quantum mechanics.

I have gone through various steps which are usually left out, such as demonstrating that we must look for a third degree polynomial. Taking those steps for granted, one can write directly the system and apply any method of solution. **I believe that there is no more direct way to find the formula for the sum of the squares, but I am ready to accepts suggestions.**

And now, lets go into something still simpler.

A child's play.

It may seem strange, but I believe that finding the formula for the sum of the squares intuitively, or, at least, graphically, is more complicated than for the case of the sums of the cubes (see for example http://dainoequinoziale.it/resources/scienze/matematica/sommacubi.pdf and use Google Translate if you are not familiar with the Italian language).

Let's now make a practical example, taking n=4.
We first construct three irregular tetrahedrons of equal shape, each of which has 4 floors, composed of 16 (= 4 x 4) elementary cubes on the first floor, 9 (= 3x 3) on the second floor; 4 (= 2x2) on the third floor, 1 on the top floor. As you can see, each tetrahedron is made up of S(4) elementary cubes, as it is equal to the sum of the squares of the numbers 1 to 4, i.e. 30 = 1 + 4 + 9 + 16. The smallest fantasy effort that I will require will be that you accept that the method we will adopt can be extended to n floors, that is, to the sum of the squares from 1 to n.

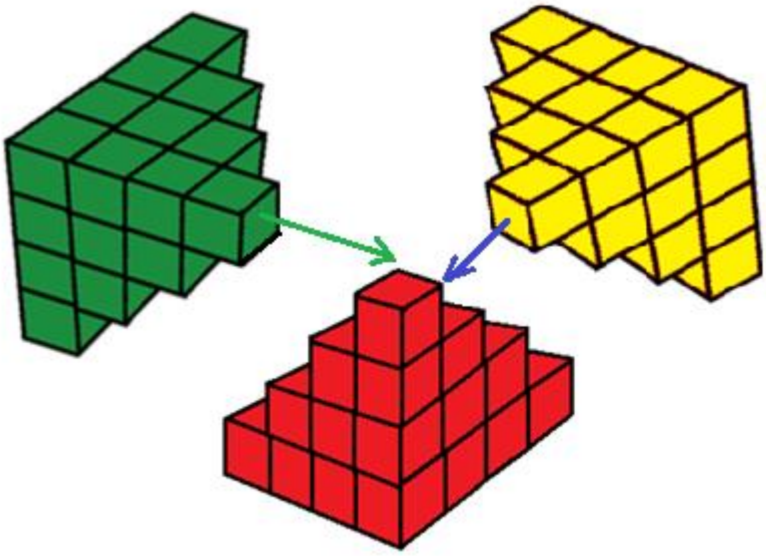

Fig.1

Let’s now fit the three tetrahedra in such a way that the various floors of the resulting “quasi-parallelepiped” are composed as shown in the following figure. **Note that these are not regular tetrahedra, and the result is not a parallelepiped.**

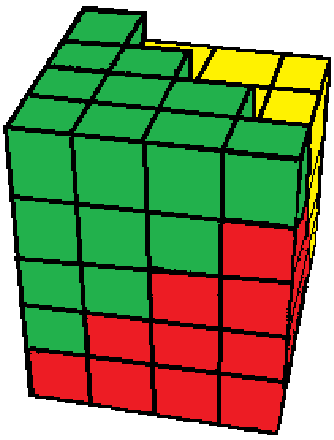

Fig.2

In fact the top floor is only half occupied by the green cubes, as the yellow tetrahedron does not have a fifth floor. Floor by floor, the arrangements looks as follows:

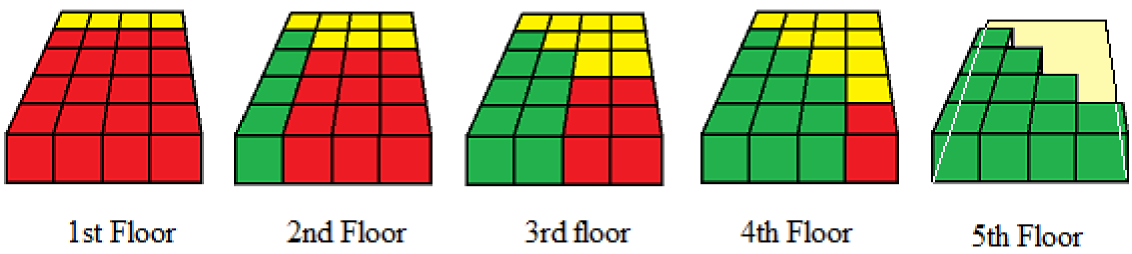

Fig.3

Fig.3

Now the polyhedron in Fig. 2 is not a parallelepiped, because the fifth floor is incomplete. To make it a parallelepiped, however, there is a simple trick. The ten green cubes on the top floor are halved vertically. Ten half-cubes will remain where they are, the other ten half cubes will be moved to the empty space. In short, the loft with terrace is given up, but there will be lower attics. At this point we have our final figure 4, which, on the top floor, has 20 rooms of height 1/2, while all the other floors have cubic rooms of height 1. The red broken line indicates where the old cubes ended and where the new half cubes are arranged.

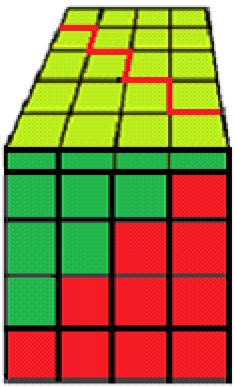

Fig.4

Parallelepiped with refurbished fifth floor

The volume of the parallelepiped remains equal to three times the sum of the squares from 1 to 4, i.e. from 1 to n, but the three sides are: n (here 4), n +1 (here 5) and the height n + 1/2 (here 4+1/2).

We therefore have:

$3\ S(n) = n\ (n + 1)\ (n + 1/2) = 1/2\ [n\ (n + 1)\ (2n + 1)$

$S(n) = (n\ (n + 1)\ (2n + 1))\ /\ 6$

Which is directly the formula preferred in Quantum Mechanics.

NOTE.

I am grateful for the formula I have presented to the
https://www.youtube.com/watch?v=aXbT37IlyZQ

YouTube site, although anonymous, which indicates that the method was probably invented elsewhere.

A more complete explanation is given in
https://en.wikipedia.org/wiki/Square_pyramidal_number

Anyway, to whomever invented it, *Chapeau !*, as the French say.