

A new approach to prime numbers and why Golbach's conjecture is true

Pedro Caceres

Longmeadow, MA- March 20th, 2017

Email: pcaceres@comcast.net

Phone: +1 (763) 412-8915

Abstract: A prime number (or a prime) is a natural number greater than 1 that has no positive divisors other than 1 and itself.

The crucial importance of prime numbers to number theory and mathematics in general stems from the fundamental theorem of arithmetic, which states that every integer larger than 1 can be written as a product of one or more primes in a way that is unique except for the order of the prime factors. Primes can thus be considered the “basic building blocks”, the atoms, of the natural numbers.

There are infinitely many primes, as demonstrated by Euclid around 300 BC. There is no known simple formula that separates prime numbers from composite numbers. However, the distribution of primes, that is to say, the statistical behavior of primes in the large, can be modelled. The first result in that direction is the prime number theorem, proven at the end of the 19th century, which says that the probability that a given, randomly chosen number n is prime is inversely proportional to its number of digits, or to the logarithm of n .

The way to build the sequence of prime numbers uses sieves, an algorithm yielding all primes up to a given limit, using only trial division method which consists of dividing \underline{n} by each integer \underline{m} that is greater than 1 and less than or equal to the square root of \underline{n} . If the result of any of these divisions is an integer, then \underline{n} is not a prime, otherwise it is a prime.

This paper introduces a new way to approach prime numbers, called the DNA-prime structure because of its intertwined nature, and a new process to create the sequence of primes without direct division or multiplication, which will allow us to introduce a new primality test, and a new factorization algorithm.

As a consequence of the DNA-prime structure, we will be able to provide a potential proof of Golbach's conjecture.

A. INTRODUCTION

1. Prime sequence

A prime number (or a prime) is a natural number greater than 1 that has no positive divisors other than 1 and itself. The sequence of prime numbers is infinite, is composed only by odd numbers and there is no known formula to generate it.

The first 25 prime numbers are given by:

2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61, 67, 71, 73, 79, ...

The infinite sequence has been studied in many different ways:

2. Differences between consecutive primes

The differences between two of these consecutive primes is calculated to be:

2	3	5	7	11	13	19	23	29	31	37	41	47	53	59	61	67	71	73	79	83	89	97
	1	2	2	4	2	6	4	6	2	6	4	6	6	6	2	6	4	2	6	4	6	8

With:

$$g_n = p_n - p_{n-1}$$

Verifying that:

$$\lim_{n \rightarrow \infty} g_n = \infty$$

And:

$$\lim_{n \rightarrow \infty} \frac{g_n}{p_n} = 0$$

The differences between primes are increasing and the prime number theorem proved that these gaps grows with the logarithm of n. The function is neither multiplicative nor additive.

As of March 2017 the largest known prime gap with identified probable prime gap ends has length 5103138, with 216849-digit probable primes found by Robert W. Smith.[3] This gap has merit M=10.2203. The largest known prime gap with identified proven primes as gap ends has length 1113106, with 18662-digit primes found by P. Cami, M. Jansen and J. K. Andersen [4][5]

3. Ratios between consecutive primes

The ratios between two consecutive primes is given by:

2	3	5	7	11	13	19	23	29	31	37	41	47	53	59
	1.50	1.67	1.40	1.57	1.18	1.46	1.21	1.26	1.07	1.19	1.11	1.15	1.13	1.11

These ratios are decreasing with:

$$\lim_{n \rightarrow \infty} \frac{p_n}{p_{n-1}} = 1$$

The gaps are not consistently decreasing and important research has been done on the limits of those gaps. This research is related to the counting of the number of primes less than a given number.

4. Known algebraic expressions for primes

The following expression by Euler connects all prime numbers with the Riemann Zeta function over all natural numbers:

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_{p \text{ prime}} \frac{1}{1 - \frac{1}{p^s}}$$

One of the most important advance in the study of Prime numbers was the paper by Bernhard Riemann in November 1859 called “Ueber die Anzahl der Primzahlen unter einer gegebenen Grösse” (On the number of primes less than a given quantity) [2].

In this paper, Riemann gave a formula for the number of primes less than x in terms the integral of 1/log(x) and the roots (zeros) of the zeta function in the complex plane, defined by:

$$\zeta(z) = \sum_{n=1}^{\infty} \frac{1}{n^z}$$

In this paper, Riemann formulated his famous hypothesis that all zeros of $\zeta(z)$ have $\text{Re}(z)=1/2$.

In a paper not published yet [12] “An Engineer’s approach to the Riemann Hypothesis and why it is true” (Feb 2017) I formulated that the following expressions that are valid for all the non-trivial zeros of Riemann $\zeta(z)$:

1. Any zero of $\zeta(z)$ with $z=\alpha+i\beta$ meet these two conditions:

(condition 1) $\alpha=1/2$

(condition 2) Calculating β :

If $S = \frac{1}{[\beta^2 + 1/4]}$ then for $n=1/S \rightarrow$

$$\left(\sum_{k=1}^n \sum_{j \neq k}^n k^{-\frac{1}{2}} * j^{-\frac{1}{2}} * \cos\left(\beta \left(\ln\left(\frac{k}{j}\right)\right)\right) \right) = 0$$

This is a finite sum, with $n \in [1, \frac{1}{S}]$

2. All zeros of $\zeta(z)$ are related through the following algebraic expression where $z_1=\alpha_1+i\beta_1$ and $z_2=\alpha_2+i\beta_2$ are non-trivial zeros of Riemann $\zeta(z)$:

$$\frac{n}{[\beta_2^2 + (1-\alpha)^2]} - \sum_{k=1}^n \sum_{j \neq k}^n k^{-1/2} * j^{-1/2} * \cos(\beta_2 \left(\ln\left(\frac{k}{j}\right)\right)) =$$

$$\frac{n}{[\beta_1^2 + (1-\alpha)^2]} - \sum_{k=1}^n \sum_{j \neq k}^n k^{-1/2} * j^{-1/2} * \cos(\beta_1 (\ln(\frac{k}{j})))$$

when $n \rightarrow \infty$.

3. The harmonic function H_n can be expressed in infinite ways as a function of each non-trivial zeros of Riemann $\zeta(z)$:

$$H_n = \frac{n}{[\beta^2 + (1-\alpha)^2]} - \sum_{k=1}^n \sum_{j \neq k}^n k^{-1/2} * j^{-1/2} * \cos(\beta (\ln(\frac{k}{j}))) \quad \text{when } n \rightarrow \infty$$

4. The Imaginary part of the non-trivial zeros of the Riemann $\zeta(z)$ is equal to the zeros of the following function in the Real plane:

$$P(n, \beta) = \sum_{k=1}^n \sum_{j \neq k}^n k^{-\frac{1}{2}} * j^{-\frac{1}{2}} * \cos\left(\beta \left(\ln\left(\frac{k}{j}\right)\right)\right)$$

where $1/2+i\beta$ would be a non-trivial zeros of Riemann $\zeta(z)$.

In the same paper, the author also provided a potential proof of the Riemann Hypothesis.

5. Number of primes less than a given number

Let's call $\pi(n)$ the number of primes less than n . The prime number theorem says that:

$$\lim_{n \rightarrow \infty} \pi(n)/(n/\ln n) = 1$$

Which can be written as:

$$\lim_{n \rightarrow \infty} \pi(n)/li(n) = 1$$

Where:

$$li(n) = \int_0^n \frac{dt}{t}$$

The following table shows the results of this approximation [6 modified]:

x	$\pi(x)$	$\pi(x) - x / \ln x$	$li(x) - \pi(x)$	$\pi(x) / li(x)$
1E+01	4	-0.34	2.200	0.645161290
1E+02	25	3.29	5.100	0.830564784
1E+03	168	23.24	10.000	0.943820225
1E+04	1,229	143.26	17.000	0.986356340
1E+05	9,592	906.11	38.000	0.996053998
1E+06	78,498	6115.59	130.000	0.998346645
1E+07	664,579	44158.31	339.000	0.999490163
1E+08	5,761,455	332773.98	754.000	0.999869147
1E+09	50,847,534	2592591.57	1701.000	0.999966548
1E+10	455,052,511	20758029.10	3104.000	0.999993179
1E+11	4,118,054,813	169923159.33	11588.000	0.999997186

Table 1

The effort in this direction is to find more accurate approximations to $\pi(n)$. All these expressions involve complex algebraic expressions of $\ln(n)$, or the Riemann Zeta function, and $li(x)$.

As an example, the Riemann hypothesis is equivalent to a much tighter bound on the error in the estimate for $\pi(n)$. and hence to a more regular distribution of prime numbers, Specifically, [9]

$$|\pi(n) - li(x)| < \frac{1}{8\pi} \sqrt{x} \ln x \text{ for all } x > 2657$$

6. Prime factorization

In number theory, the prime factors of a positive integer are the prime numbers that divide that integer with no remainder.

If \underline{n} is divided by \underline{p} , there is a $k, r \in Z$ such that:

$$n = k * p + r$$

\underline{p} is a prime factor of \underline{n} , if and only if $r = 0$, which can also be expressed using the mod(ulo) function by:

$$n \bmod p = 0$$

Where the function mod(ulo) is defined as follows:

$$r = p - n * trunc\left(\frac{p}{n}\right)$$

The prime factorization of a positive integer is a list of the integer's prime factors, together with their multiplicities; the process of determining these factors is called integer factorization. The fundamental theorem of arithmetic says that every positive integer has a single unique prime factorization.[7]

To find prime factors requires to know if a number is a prime. This type of test is called primality test. Among other fields of mathematics, it is used extensively in cryptography.

The RSA codes in cryptography consists of very large composite numbers that have exactly two prime factors. These numbers are called semiprimes. Finding those two factors require very complex algorithms as the numbers are composed by two prime numbers of more than one hundred digits. As an example:

RSA-220 =

260138526203405784941654048610197513508038915719776718321197768109445641
817966676608593121306582577250631562886676970448070001811149711863002112
487928199487482066070131066586646083327982803560379205391980139946496955
261

FACTOR 1 of RSA-220 =

686365641226756627438237149928843780013084223997916484462124499332154106
14414642667938213644208420192054999687

FACTOR 2 of RSA-220 =

329290743948634981204930154921293529191645519653623395246268605116929034
93094652463337824866390738191765712603

The simple factorization method is the trial division method which consists in dividing sequentially by all known primes until we find a factor. Then we reduce the number by the factor and start again. This method is unpractical for large primes.

The fastest-known fully proven deterministic algorithm is the Pollard-Strassen method (Pomerance 1982; Hardy et al. 1990). [8]

Wolfram Math World mentions the following list of factorization methods: [9]

- Brent's Factorization Method,
- Class Group Factorization Method,
- Continued Fraction Factorization Algorithm,
- Direct Search Factorization,
- Dixon's Factorization Method,
- Elliptic Curve Factorization Method,
- Euler's Factorization Method,
- Excludent Factorization Method,
- Fermat's Factorization Method,
- Legendre's Factorization Method,
- Number Field Sieve,
- Pollard p-1 Factorization Method,
- Pollard rho Factorization Algorithm,

- Quadratic Sieve,
- Trial Division,
- Veryprime,
- Williams p+1 Factorization Method

B. A new approach to Prime Numbers. Defining the DNA-Prime Sequences P+ and P-

So far, the sequence of primes is a mystery. We have been able to find some patterns, predict the number of primes, factor very large integers but we are still in the same position as Eratosthenes was more than 2000 years ago, when he introduced his sieve to identify prime numbers.

In this paper, we want to introduce a new way to look at primes and develop a new primality test and factorization method from it.

One characteristic that all primes greater than 3 have is that they can be expressed with one of the two following expressions:

$$p = 6k_n + 1 \quad k_n \in N \quad (1)$$

$$p = 6k_m - 1 \quad k_m \in N \quad (2)$$

In theory, if we knew the sequences k_n and k_m we would know all primes.

In this paper, we are going to provide a unique formulation for the two “generator” series k_n and k_m .

As we can see in the following table:

P+	P-		
6kn+1	6km-1	kn	km
	5	1	
7			1
	11	2	
13			2
	17	3	
19			3
	23	4	
	29		5
31		5	
37			6
	41	7	
43			7
	47	8	
	53		9
	59	10	
61			10

Table 2

The prime numbers belong to either P⁺ or P⁻ series.

It is amazing that all primes can be generated from the numbers 1,2 ,3, as:

$$P^+ \text{ series} \quad p = 1+2*3*k_n$$

$$P^- \text{ series} \quad p = -1+2*3*k_m$$

I will call these two sequences the DNA-Prime Sequences as it resembles the intertwined DNA helix.

C. Characteristics of the DNA-Prime Sequences

a. The difference between two primes in either sequence P⁺ and P⁻ is a multiple of 6. In the following table we show the two DNA-prime series, the difference between two consecutive elements of the series, the difference divided by 6 and the cumulative difference divided by 6.

We must observe that the cumulative difference from any element of the series and the first element, that we will call R_n for the P⁺ series, and R_m for the P⁻ series, is equal to the K generator minus 1. This key fact will help us formulate a way to generate the prime sequence.

P+					P-				
Pn+	Kn	P(n)-P(n-1)	(P(n)-P(n-1))/6	Rn	Pm-	Km	P(m)-P(m-1)	(P(m)-P(m-1))/6	Rm
7	1				5	1			
13	2	6	1	1	11	2	6	1	1
19	3	6	1	2	17	3	6	1	2
31	5	12	2	4	23	4	6	1	3
37	6	6	1	5	29	5	6	1	4
43	7	6	1	6	41	7	12	2	6
61	10	18	3	9	47	8	6	1	7
67	11	6	1	10	53	9	6	1	8
73	12	6	1	11	59	10	6	1	9
79	13	6	1	12	71	12	12	2	11
97	16	18	3	15	83	14	12	2	13
103	17	6	1	16	89	15	6	1	14
109	18	6	1	17	101	17	12	2	16
127	21	18	3	20	107	18	6	1	17
139	23	12	2	22	113	19	6	1	18
151	25	12	2	24	131	22	18	3	21
157	26	6	1	25	137	23	6	1	22
163	27	6	1	26	149	25	12	2	24
181	30	18	3	29	167	28	18	3	27
193	32	12	2	31	173	29	6	1	28

Table 3

We can see in the chart that:

$$\begin{array}{ll}
 P^+ \text{ series} & P(n)-P(n-1) \bmod 6 = 0 \\
 P^- \text{ series} & P(m)-P(m-1) \bmod 6 = 0
 \end{array}$$

b. The difference between any two primes in either series is given by:

$$\begin{array}{llll}
 P^+ \text{ series} & \text{if } p_1=6*k_1+1 & \text{and} & p_2=6*k_2+1 \\
 P^- \text{ series} & \text{if } p_1=6*k_1-1 & \text{and} & p_2=6*k_2-1
 \end{array}$$

$$\text{Then } p_2-p_1 = 6 * (k_2-k_1)$$

c. The difference between any prime in either sequence P⁺ and P and the first one in the series is a multiple of the generators:

$$\begin{array}{ll}
 P^+ \text{ series} & P_n = 7 + 6*R_n \\
 P^- \text{ series} & P_m = 5 + 6*R_m
 \end{array}$$

$$\text{Where } R_n = k_n-1 \quad \text{and} \quad R_m = k_m-1$$

If we can find a formula for the generators, then we will have a formula for the primes. Let's take a look at the R_n and R_m sequences eliminating (in color) all those that don't generate a prime using previous formulas:

R_n

1	2	3	4	5	6	7	8	9	10
11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30
31	32	33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48	49	50
51	52	53	54	55	56	57	58	59	60
61	62	63	64	65	66	67	68	69	70
71	72	73	74	75	76	77	78	79	80
81	82	83	84	85	86	87	88	89	90
91	92	93	94	95	96	97	98	99	100
101	102	103	104	105	106	107	108	109	110
111	112	113	114	115	116	117	118	119	120
121	122	123	124	125	126	127	128	129	130
131	132	133	134	135	136	137	138	139	140
141	142	143	144	145	146	147	148	149	150
151	152	153	154	155	156	157	158	159	160
161	162	163	164	165	166	167	168	169	170
171	172	173	174	175	176	177	178	179	180
181	182	183	184	185	186	187	188	189	190

Table 4

R_m

1	2	3	4	5	6	7	8	9	10
11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30
31	32	33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48	49	50
51	52	53	54	55	56	57	58	59	60
61	62	63	64	65	66	67	68	69	70
71	72	73	74	75	76	77	78	79	80
81	82	83	84	85	86	87	88	89	90
91	92	93	94	95	96	97	98	99	100
101	102	103	104	105	106	107	108	109	110
111	112	113	114	115	116	117	118	119	120
121	122	123	124	125	126	127	128	129	130
131	132	133	134	135	136	137	138	139	140
141	142	143	144	145	146	147	148	149	150
151	152	153	154	155	156	157	158	159	160
161	162	163	164	165	166	167	168	169	170
171	172	173	174	175	176	177	178	179	180
181	182	183	184	185	186	187	188	189	190

Table 5

D. The prime series can be formulated algebraically

From the tables above and testing many potential combinations, we have concluded that the sequence of DNA-primes and their generators R_n and R_m can be formulated algebraically as follows:

P⁺ series	$P_n = 7 + 6 \cdot R_n$	
P⁺ series	$R_n \neq x + (6x + 7) \cdot y$	$x > 0, y > 1 \in N$
	$R_n \neq -x + (6x - 7) \cdot y$	$x > 1, y > 1 \in N$
P⁻ series	$P_m = 5 + 6 \cdot R_m$	
P⁺ series	$R_m \neq x + (6x + 5) \cdot y$	$x > 0, y > 1 \in N$
	$R_m \neq -(x + 1) + (6x + 1) \cdot y$	$x > 0, y > 1 \in N$

These expressions are equivalent to:

P ⁺ series	P _n = 1 + 6 · K _n	with K _n = R _n + 1
-----------------------	---	--

$$P \text{ series} \quad P_m = -1 + 6 \cdot K_m \quad \text{with } K_m = R_m + 1$$

The first numbers in the generator series are R_n and R_m :

$R_n = 1, 2, 4, 5, 6, 9, 10, 11, 12, 15, 16, 17, 20, 22, 24, 25, 26, 29, 31, 32, 34, 36, 37, 39, 44, 45, 46, 50, 51, 54, 55, 57, 60, 61, 62, 65, \dots$

$R_m = 1, 2, 3, 4, 6, 7, 8, 9, 11, 13, 14, 16, 17, 18, 21, 22, 24, 27, 28, 29, 31, 32, 37, 38, 39, 41, 42, 43, 44, 46, 48, 51, 52, 57, 58, 59, 63, \dots$

And the primes generated by these generators:

Primes generated = [2, 3, 5, 7], 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61, 67, 71, 73, 79, 83, 89, 97, 101, 103, 107, 109, 113, 127, 131, 137, 139, 149, ...

We prepared a simple code to run the prime generation and we this is a test for the prime numbers less than $n=1,000,000$

```

PEDRO CACERES - pcaceres@comcast.net - +1 (763) 412-8915
Version 03/14/2017
-----
GENERATING PRIME SERIES Start @ 2017-03-21 11:44:11
-----
Number of Primes      = 78498
Number of Prime Pairs = 8168 %Pi(x)= 10.405360646131111
-----
My series generated   @ 2017-03-21 11:49:27
---> GENERATED Prime list is correct
-----
Last prime position: 78498 -> 999983
-----
GENERATING PRIME SERIES Ends @ 2017-03-21 11:49:27
-----

```

E. The gaps between elements of the DNA-prime series

The gaps between elements of R_n and R_m can be described using the Prime Number Theorem as the DNA-prime series and the prime series have a bijective relationship.

We can affirm that:

$$R_n - R_{n-1} \sim 1/6 * \ln(6 * R_n + 7) < R_{n-1} < R_n$$

$$R_m - R_{m-1} \sim 1/6 * \ln(6 * R_m + 5) < R_{m-1} < R_m$$

F. DNA-Prime twins

The definition of the two DNA-prime series implies that any S_{2n} such that $S_{2n} = R_n = R_m$ will create a pair of twin primes.

The list of the first S_{2n} is:

$$S_{2n} = [1, 2, 4, 6, 9, 11, 16, 17, 22, 24, 29, 31, 32, 37, 39, 44, 46, \dots]$$

And the pairs generated:

Twin Primes from $S_{2n} =$	1	= [11 , 13]
Twin Primes from $S_{2n} =$	2	= [17 , 19]
Twin Primes from $S_{2n} =$	4	= [29 , 31]
Twin Primes from $S_{2n} =$	6	= [41 , 43]
Twin Primes from $S_{2n} =$	9	= [59 , 61]
Twin Primes from $S_{2n} =$	11	= [71 , 73]
Twin Primes from $S_{2n} =$	16	= [101 , 103]
Twin Primes from $S_{2n} =$	17	= [107 , 109]
Twin Primes from $S_{2n} =$	22	= [137 , 139]
Twin Primes from $S_{2n} =$	24	= [149 , 151]
Twin Primes from $S_{2n} =$	29	= [179 , 181]
Twin Primes from $S_{2n} =$	31	= [191 , 193]
Twin Primes from $S_{2n} =$	32	= [197 , 199]
Twin Primes from $S_{2n} =$	37	= [227 , 229]
Twin Primes from $S_{2n} =$	39	= [239 , 241]
Twin Primes from $S_{2n} =$	44	= [269 , 271]
Twin Primes from $S_{2n} =$	46	= [281 , 283]

We must add [3,5] and [5,7] to complete the series.

The series S_{2n} has infinite terms as R_n and R_m are also infinite and the condition for $R_n = R_m$ does not have a maximum limit. The generation of R_n and R_m is done over all $x, y \in N$. The proof will only require to define a x, y that makes all 4 conditions of the DNA-series not equal. The length of the gap among twin prime sets can be calculated from the difference between elements in S_{2n} .

The Brun's theorem [11] proved that the sum of the reciprocals of these twin primes converges to a finite value known as Brun's constant, usually denoted by B_2 . In 2002 Pascal Sebah and Patrick Demichel used all twin primes up to 10^{16} to give the estimate:

$$B_2 = \sum_{p,p+2 \text{ pairs}} \left(\frac{1}{p} + \frac{1}{p+2} \right) = 1.902160583104 \quad [6]$$

We can express this equation with the DNA-prime expression as:

$$B_2 = \sum_{0,R=Rn=Rm} \left(\frac{1}{(6*R+7)} + \frac{1}{(6*R+5)} \right) = 1.902160583104 \quad [6]$$

Adding $(1/3+1/5)$.

Using again the definition of the two DNA-prime series we can calculate other partial consecutive sequences of primes S_{kn} .

For instances, series of 4 consecutive primes will require that S_{4n} and S_{4n+1} to be both in R_n , and R_m .

The list of the first S_{4n} is:

$$S_{4n} = [1, 16, 31, 136, 246, 311, 346, 541, 576, 941, 1571, \dots]$$

With:

$S_{4n} =$	1	=	[11 , 13 , 17 , 19]
$S_{4n} =$	16	=	[101 , 103 , 107 , 109]
$S_{4n} =$	31	=	[191 , 193 , 197 , 199]
$S_{4n} =$	136	=	[821 , 823 , 827 , 829]
$S_{4n} =$	246	=	[1481 , 1483 , 1487 , 1489]
$S_{4n} =$	311	=	[1871 , 1873 , 1877 , 1879]
$S_{4n} =$	346	=	[2081 , 2083 , 2087 , 2089]
$S_{4n} =$	541	=	[3251 , 3253 , 3257 , 3259]
$S_{4n} =$	576	=	[3461 , 3463 , 3467 , 3469]
$S_{4n} =$	941	=	[5651 , 5653 , 5657 , 5659]
$S_{4n} =$	1571	=	[9431 , 9433 , 9437 , 9439]

We must add [5 , 7 , 11 , 13] to complete the series,

There is also a Brun's constant for these prime quadruplets with $B_4 = 0.87058838\dots$ Wolf derived an estimate for the Brun-type sums B_n of $4/n$.

I have not found any elements for sets for S_{6n} and above yet.

G. DNA-Prime primality test

For a number N to be prime, the following conditions must be met:

- a) If $(N-1) \bmod 6 \neq 0$ and $(N+1) \bmod 6 \neq 0$ the number is not prime
- b) If $(N-1) \bmod 6 = 0$ then $N=6 \cdot k_n+1$ and $R_n=K_n-1$
- CONDITION 1 $(R_n - 7s) \bmod (6s + 1) \neq 0$ for $s \in N < k_n$
- CONDITION 2 $(R_n + 7s) \bmod (6s - 1) \neq 0$ for $s \in N < k_n$

If $s=1$ or $s=k_n$ then N is Prime.

- c) If $(N+1) \bmod 6 = 0$ then $N=6 \cdot k_m+1$ and $R_m=K_m-1$
- CONDITION 3 $(R_m - (s - 1)) \bmod (6s - 1) \neq 0$ for $s \in N < k_m$
- CONDITION 4 $(R_m - 5s) \bmod (6s + 1) \neq 0$ for $s \in N < k_m$

If $s=1$ or $s=k_m$ then N is Prime.

Examples:

	Kn	Km	Rn	Rm	s	Condition	
4,489	748.00	748.33	747.00		11	1	No prime
6,839	1139.67	1140.00		1139.00	163	3	No prime
9,973	1662.00	1662.33	1661.00		1662		Prime

The maximum values of (s) for each condition can be formulated as a function of N to optimize the algorithm.

H. DNA-Prime factorization process

For any given number N, the factors can be calculated with the following method:

1. Check if the number is divisible by 2,3,5,7
 - a. If yes, divide N by factors and continue with $N^*=N/\text{factors}$ ($2^a, 3^b, 5^c, 7^d$)
2. If $N^* > 7$, Determine if N^* belongs to any DNA-Prime Sequence P^+ or P^-
 - a. If $N \in P^+$ then calculate $k_n=(N-1)/6$ and $R_n=k_n-1$ and check Condition 1
 - i. If Condition 1 is not met at $s < k_n$ then $(6 \cdot s - 1)$ is a prime factor of N^* . Divide N^* by the factor. $N^{**}=N^*/\text{factor}$
 - ii. If condition 1 is met the check Condition 2

1. If Condition 2 is not met at $s < k_n$ then $(6*s+1)$ is a prime factor of N^* . Divide N^* by the factor. $N^{**}=N^*/\text{factor}$
 2. If condition 2 is met, N^* is prime
- b. If $N \in P^-$ then then calculate $k_m=(N+1)/6$ and $R_m=k_m-1$ and check Condition 3
- i. If Condition 3 is not met at any $s < k_m$ then $(6*s-1)$ is a prime factor of N^* . Divide N^* by the factor. $N^{**}=N^*/\text{factor}$
 - ii. If condition 3 is met then Check Condition 4
 1. If Condition 4 is not met at $s < k_m$ then $(6*s+1)$ is a prime factor of N^* . Divide N^* by the factor. $N^{**}=N^*/\text{factor}$
 2. If condition 4 is met, N^* is prime
- c. If N^* is not prime run process for new $N^{**}=N^*/\text{factor}$

Some examples of factorization with computer time:

1. Factoring N = 1234567890 1234567890 showing all the steps:

```

PEDRO CACERES - pcaceres@comcast.net - +1 (763) 412-8915
FACTORIZA Version 03/14/2017
-----
FACTORIZA Start @ 2017-03-20 19:03:22
-----
N= 12345678901234567890
-----
PARTIAL N= 6172839450617283945 ->FACTORS+ [2]
-----
PARTIAL N= 2057613150205761315 ->FACTORS+ [2, 3]
-----
PARTIAL N= 411522630041152263 ->FACTORS+ [2, 3, 5]
-----
N= 411522630041152263 is NOT P+ or P-
-----
PARTIAL N= 137174210013717421 ->FACTORS+ [2, 3, 5, 3]
-----
N= 137174210013717421 is in Series: P+
-----
N= 137174210013717421
k= 22862368335619570
r= 22862368335619569
-----
N: 137174210013717421 -> Failed Condition 1+ @ s= 590
-----
PARTIAL N= 38738833666681 ->FACTORS+ [2, 3, 5, 3, 3541]
-----
N= 38738833666681 is in Series: P+
-----
N= 38738833666681
k= 6456472277780
r= 6456472277779
-----
N: 38738833666681 -> Failed Condition 1+ @ s= 601
-----

```

```

PARTIAL N= 10739903983 ->FACTORS+ [2, 3, 5, 3, 3541, 3607]
-----
N= 10739903983 is in Series: P+
-----
N= 10739903983
k= 1789983997
r= 1789983996
-----
N: 10739903983 -> Failed Condition 1+ @ s= 4660
-----
PARTIAL N= 384103 ->FACTORS+ [2, 3, 5, 3, 3541, 3607, 27961]
-----
N= 384103 is in Series: P+
-----
N= 384103
k= 64017
r= 64016
-----
N: 384103 -> Failed Condition 2+ @ s= 17
-----
PARTIAL N= 3803 ->FACTORS+ [2, 3, 5, 3, 3541, 3607, 27961, 101]
-----
N= 3803 is in Series: P-
-----
N= 3803
k= 634
r= 633
-----
PARTIAL N= 3803 ->FACTORS+ [2, 3, 5, 3, 3541, 3607, 27961, 101]
-----
N= 12345678901234567890 --> Factors= 2*3^2*5*101*3541*3607
-----
FACTORIZA Ends @ 2017-03-20 19:03:33

```

2. Factorization of 20 consecutive numbers without factorization steps:

PEDRO CACERES - pcaceres@comcast.net - +1 (763) 412-8915
FACTORIZA Version 03/14/2017

```

-----
FACTORIZA Start @ 2017-03-20 22:37:51
-----
N= 12345678901234567890 --> Factors= 2 3^2 5 101 3541 3607 3803 27961
N= 12345678901234567891 --> PRIME
N= 12345678901234567892 --> Factors= 2^2 3086419725308641973
N= 12345678901234567893 --> Factors= 3 14210467 289591207693
N= 12345678901234567894 --> Factors= 2 17 31 43 1189 229099455043
N= 12345678901234567895 --> Factors= 5 9577219 257813440441
N= 12345678901234567896 --> Factors= 2^3 3 5 19 389 757 7177 1830053
N= 12345678901234567897 --> Factors= 6373 1937184826805989
N= 12345678901234567898 --> Factors= 2 6172839450617283949
N= 12345678901234567899 --> Factors= 3^2 11261 73039 1667787409
N= 12345678901234567900 --> Factors= 2^2 5^2 11 12517 22147 40486211
N= 12345678901234567901 --> Factors= 314312807 39278319643
N= 12345678901234567902 --> Factors= 2 3 13^2 12175225740862493
N= 12345678901234567903 --> Factors= 5 59 2917 10247749396943
N= 12345678901234567904 --> Factors= 2^5 10133 503779 75576521
N= 12345678901234567905 --> Factors= 3 5 4314587 190758758621
N= 12345678901234567906 --> Factors= 2 23 515629579 520498309
N= 12345678901234567907 --> Factors= 1151 1223 8770274702459
N= 12345678901234567908 --> Factors= 2^2 3^4 22727 1676593797071
N= 12345678901234567909 --> Factors= 488899381 25251983089

```


I. Use of DNA-Prime to prove Golbach’s conjecture

The Golbach’s conjecture says that every even integer greater than 2 can be expressed as the sum of two primes.[10]

From the definition of the two DNA-Prime sequences we know that any prime can be expressed as:

$$p = 6k_n + 1 \quad k_n \in N$$

$$p = 6k_m - 1 \quad k_m \in N$$

The addition of two odd prime numbers will always be even.

If $N=2q$ is any even number, for it to be the addition of two primes, the following needs to be true:

$$N = 2q = p_1 + p_2$$

The possible combinations of odd numbers that added together amount to N are:

<u>a</u>	<u>b</u>	<u>a+b=N</u>
1	N-1	N
3	N-3	N
5	N-5	N
...
N-1	1	N

To illustrate the problem, let’s build a simple table for $N=18$

<u>N</u>	<u>p1</u>	<u>p2</u>
18	1	17
	3	15
	5	13
	7	11
	9	9
	11	7
	13	5
	15	3

We can see that there are $N/2$ combinations of two odd numbers that add up to N .

We can also see that there are 2 combinations involving 1 and 1 is not a prime.

The option $N/2+N/2=N$ does not involve addition of primes and we can disregard it.

Of the remaining combinations, they repeat themselves due to the commutative property of the addition in N .

So, the net number of potential valid combinations of two odd numbers with one of them at least being prime is:

$$\frac{\frac{N}{2} - 3}{2}$$

If p is prime, based on the Prime number theorem, we can see that for $N > 76$ the number of combinations is larger than the number of primes $< N$ as:

$$\frac{\frac{N}{2} - 3}{2} > N / \ln N \quad \text{for } N > 76$$

So, the number of primes that meet Golbach's conjecture for any even number N are proportionally less than the number of combinations of odd numbers as N grows.

We know that if p_1 and p_2 are primes, we can use the DNA-Prime series to say:

$$p_1 = 6 * k_1 \pm 1$$

$$p_2 = 6 * k_2 \pm 1$$

And there are three possibilities:

$$N = p_1 + p_2 = 6 * (k_1 + k_2) - 2$$

$$N = p_1 + p_2 = 6 * (k_1 + k_2)$$

$$N = p_1 + p_2 = 6 * (k_1 + k_2) + 2$$

Based on this, and assuming that p_1 and p_2 exist, we can affirm that:

- If $N \bmod 6=0$ N is the addition of a $p_1 \in P^+$ and $p_2 \in P^-$
 If $(N-2) \bmod 6=0$ N is the addition of a $p_1 \in P^+$ and $p_2 \in P^+$
 If $(N+2) \bmod 6=0$ N is the addition of a $p_1 \in P^-$ and $p_2 \in P^-$

given that for any even number, there is a $q \in N$ such that $N=2q$ and the previous expressions are equivalent to:

$$q \bmod 3 = 0$$

or $(q-1) \bmod 3 = 0$

or $(q+1) \bmod 3 = 0$

Which is obviously true as for any 3 consecutive numbers $(q-1)$, q , $(q+1)$, one of them must necessarily be divisible by 3.

The three possible combinations of primes mentioned earlier can be also reformulated as follows:

If $q \bmod 3 = 0$ $\frac{q}{3} = k_1 + k_2$ $p_1 = 6k_1 + 1$ and $p_2 = 6k_2 - 1$
 If $(q-1) \bmod 3 = 0$ $\frac{q-1}{3} = k_1 + k_2$ $p_1 = 6k_1 + 1$ and $p_2 = 6k_2 + 1$
 If $(q+1) \bmod 3 = 0$ $\frac{q+1}{3} = k_1 + k_2$ $p_1 = 6k_1 - 1$ and $p_2 = 6k_2 - 1$

As examples of these expressions:

				Potential k ₁	Potential k ₂	Primes	
N=54	q=27	q=3*9	q/3=9	2	7	13	41
N=64	q=32	q=3*11-1	(q+1)/3=11	3	8	17	47
N=68	q=34	q=3*11+1	(q-1)/3=11	5	6	31	37

To prove Golbach's conjecture we must prove that for any $n \in N$ we can find combinations of R_n and R_m in the DNA-Prime generator series such that:

$$q = k_n + k_m = R_n + R_m - 2$$

$$\text{or } q = k_{n1} + k_{n2} = R_n^1 + R_n^2 - 2$$

$$\text{or } q = k_{m1} + k_{m2} = R_m^1 + R_m^2 - 2$$

In other words, that for any $q \in N$, we can find two elements of R_n , or two elements of R_m , or one element of R_n and one element of R_m , that add up to q , as all even numbers are of the form: $N=2q$, with $q \in N$

To prove it, we are going to use an induction proof.

We will define a condition that is observable and met for a certain $k=k^*$, we will assume that the condition is met at $k=n-1$ and then we will prove that this means the condition is also true at $k=n$ for any element n of the generator series.

Let's observe that in the following chart of (R_m+R_m) , the square $R_m \times R_m$ contains at least all naturals up to R_m .

Rm x Rm table															
	1	2	3	4	6	7	8	9	11	13	14	16	17	18	
1	2	3	4	5	7	8	9	10	12	14	15	17	18	19	
2	3	4	5	6	8	9	10	11	13	15	16	18	19	20	
3	4	5	6	7	9	10	11	12	14	16	17	19	20	21	
4	5	6	7	8	10	11	12	13	15	17	18	20	21	22	
6	7	8	9	10	12	13	14	15	17	19	20	22	23	24	
7	8	9	10	11	13	14	15	16	18	20	21	23	24	25	
8	9	10	11	12	14	15	16	17	19	21	22	24	25	26	
9	10	11	12	13	15	16	17	18	20	22	23	25	26	27	
11	12	13	14	15	17	18	19	20	22	24	25	27	28	29	
13	14	15	16	17	19	20	21	22	24	26	27	29	30	31	
14	15	16	17	18	20	21	22	23	25	27	28	30	31	32	
16	17	18	19	20	22	23	24	25	27	29	30	32	33	34	
17	18	19	20	21	23	24	25	26	28	30	31	33	34	35	
18	19	20	21	22	24	25	26	27	29	31	32	34	35	36	

Table 6

For example, the square for $(R_{11} \times R_{11})=(11 \times 11)$ contains up to the natural number up to 11, i.e. contains $[1], 2, 3, 4, 5, 6, 7, 8, 9, 10, 11$

We could use any other cardinal to observe that this true.

Let's assume now that the condition is true for R_{m-1} , which means that the square $R_{m-1} \times R_{m-1}$ contains all naturals up to R_{m-1} generated by the addition of two given R_m^1 and R_m^2 both $< R_{m-1}$ and let's prove that the condition is met for the square $R_m \times R_m$, which means that the square $R_m \times R_m$ must contain all naturals up to R_m .

The set of natural numbers between R_{m-1} and R_m are, by definition of the matrix

$R_m \times R_m$:

$$D_n = \{ R_m - R_{m-1} \} = \{ R_{m-1} + 1, R_{m-1} + 2, R_{m-1} + 3 \dots, R_{m-1} + (R_m - R_{m-1}) \}$$

We know that all naturals up to R_{m-1} exists and for each $n < R_{m-1}$ there are two:

$$R_{m-1}^j = \{ R_{m-1}^1 \dots R_{m-1}^j \}$$

$$R_{m-1}^k = \{ R_{m-1}^1 \dots R_{m-1}^k \}$$

Such that $n = R_{m-1}^j + R_{m-1}^k$

The assumption over R_{m-1} implies that the set of $\{ R_{m-1}^j + R_{m-1}^k \}$ generates all naturals up to R_{m-1}

$$\{ R_{m-1}^j + R_{m-1}^k \} = \{ 2, 3, 4, 5, \dots R_{m-1} \}$$

$$\text{with } R_{m-1}^j = \{ 1, 2, 3, 4, 5, 6, 8, 9, \dots R_{m-1} \}$$

If we add R_{m-1} to any of the two $R_{m-1}^j + R_{m-1}^k$ we can say that:

$$\{ R_{m-1} + R_{m-1}^j \} = \{ R_{m-1} + 1, R_{m-1} + 2 \dots, 2 * R_{m-1} \}$$

$2 * R_{m-1}$ is the last diagonal term of the defined and known matrix $R_{m-1} \times R_{m-1}$ which is contained in $R_m \times R_m$

And we know from [D] that $R_m - R_{m-1} < 2 * R_{m-1}$ so, therefore, D_n is contained in the matrix $R_m \times R_m$.

Same proof works for (R_n+R_n) and (R_n+R_m) , which proves the Golbach's conjecture.

As an example, to find the primes that add up to 180:

$N=180$ $N \bmod 6=0$ so one prime belongs to P^+ and the other to P^-
 $q=N/6=30$ and $R_n+R_m=30-2=28$

All the combinations of $R_n+R_m=28$ are:

R_n	R_m	P^+	P^-	P^++P^-
1	27	13	167	= 180
4	24	31	149	= 180
6	22	43	137	= 180
10	18	67	113	= 180
11	17	73	107	= 180
12	16	79	101	= 180
15	13	97	83	= 180
17	11	109	71	= 180
20	8	127	53	= 180
22	6	139	41	= 180
24	4	151	29	= 180
25	3	157	23	= 180
26	2	163	17	= 180

Table 7

J. Open challenges

- Can a formula for $\pi(x)$ counting all primes up to x be formulated exactly based on the 4 conditions of the DNA-prime series? To calculate this, we should be able to calculate the amount of repetitions among the four conditions.
- Can a formula for gaps between primes and twin primes be formulated based on the 4 conditions of the DNA-prime series?

Thanks
 Pedro Caceres
pcaceres@comcast.net
 +1 (763) 412-8915

REFERENCES

- (1) Computer Programs used in this proof designed and developed by Pedro J. Caceres
- (2) B. Riemann, "Ueber die Anzahl der Primzahlen unter einer gegebenen Grösse", *Monat. der Königl. Preuss. Akad. der Wissen. zu Berlin* aus der Jahre 1859 (1860), 671–680; also, *Gesammelte math. Werke und wissenschaft. Nachlass*, 2. Aufl. 1892, 145–155.
- (3) <http://trnicely.net/index.html#TPG>
- (4) Andersen, Jens Kruse. "The Top-20 Prime Gaps". Retrieved 2014-06-13.
- (5) Jump up^ A proven prime gap of 1113106
- (6) "A table of values of $\pi(x)$ ". Xavier Gourdon, Pascal Sebah, Patrick Demichel. Retrieved 2008-09-14.
- (7) Schoenfeld, Lowell (1976). "Sharper bounds for the Chebyshev functions $\theta(x)$ and $\psi(x)$. II". *Mathematics of Computation*. American Mathematical Society. 30 (134): 337–360. doi:10.2307/2005976. ISSN 0025-5718. JSTOR 2005976. MR 0457374.
- (8) Riesel, Hans (1994), *Prime numbers and computer methods for factorization*, Basel, Switzerland: Birkhäuser, ISBN 978-0-8176-3743-9
- (9) <http://mathworld.wolfram.com/PrimeFactorizationAlgorithms.html>
- (10) Weisstein, Eric W. "Goldbach Conjecture". MathWorld.
- (11) Brun, Viggo (1919). "La série $1/5+1/7+1/11+1/13+1/17+1/19+1/29+1/31+1/41+1/43+1/59+1/61+\dots$, où les dénominateurs sont nombres premiers jumeaux est convergente ou finie". *Bulletin des Sciences Mathématiques (in French)*. 43: 100–104, 124–128.
- (12) Pedro Caceres, "An Engineer's approach to the Riemann Hypothesis and why it is true", February 2017. [HTTP://vixra.org/abs/1703.0104](http://vixra.org/abs/1703.0104)