



HAL
open science

A proof of Sophie Germain primes conjecture

Marko V Jankovic

► **To cite this version:**

| Marko V Jankovic. A proof of Sophie Germain primes conjecture. 2022. hal-02169242v12

HAL Id: hal-02169242

<https://hal.science/hal-02169242v12>

Preprint submitted on 20 Feb 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Proof of the Sophie Germain Primes Conjecture

Marko V. Jankovic

Institute of Electrical Engineering “Nikola Tesla”, Belgrade, Serbia

to Judy

Abstract In this paper a proof of the existence of an infinite number of Sophie Germain primes is going to be presented. Originally very difficult problem (in observational space) has been transformed into a simpler one (in generative space) that can be solved. It will be shown that Sophie Germain primes could be obtained through two stage sieve process, and that will be used to prove that infinitely many Sophie Germain primes exists.

1 Introduction

A prime p is a Sophie Germain prime if $2p + 1$ is a prime too [1]. In that case the prime number $2p + 1$ is called safe prime. These special primes have applications in public key cryptography, pseudorandom number generation, and primality testing; see, for example [2, 3, 4]. Originally, they have been used in the investigation of cases of Fermat's last theorem [5]. It has been conjectured that infinitely many Sophie Germain primes exist, but this was unproven (see for instance, [6]).

In this paper it is going to be proved that an infinite number of Sophie Germain primes exists. The problem is addressed in generative space, which means that prime numbers are not going to be analyzed directly, but rather their representatives that are used to produce them. It will be shown that Sophie Germain primes could be generated by two stage recursive type process. This process will be compared to another two stage sieve process that leaves infinitely many numbers. Fact that sieve process that generate Sophie Germain primes leaves more numbers than the other sieve process, will be used to prove that infinitely many Sophie Germain primes exist.

Remark 1: *In this paper any infinite series in the form $c_1 * l \pm c_2$ is going to be called a thread defined by number c_1 (in literature these forms are known as linear factors – however, it seems that the term thread is probably better choice in this context). Here c_1 and c_2 are numbers that belong to*

the set of natural numbers (c_2 can also be zero and usually is smaller than c_1) and l represents an infinite series of consecutive natural numbers in the form $(1, 2, 3, \dots)$.

2 Proof of the conjecture

It is easy to check that any prime number (apart from 2 and 3, which are Sophie Germain primes) can be expressed in the form $6l + 1$ or $6s - 1$ ($l, s \in \mathbb{N}$). Having that in mind, it is easy to conclude that numbers in the form $6l + 1$ could never be Sophie Germain primes since their “safe” primes are in the form

$$2(6l + 1) + 1 = 12l + 3 = 3(4l + 1),$$

and that is a composite number divisible by 3. Hence, the prime number that can potentially be a Sophie Germain prime must be in the form $6s - 1$ (here numbers 2 and 3 are ignored, and that obviously has no impact on the analysis that follows). The safe prime will then be in the form $6(2s) - 1$. In the text that follows, a number in the form $6l + 1$ is denoted with mpl , while a number in the form $6s - 1$ is denoted with mps ($l, s \in \mathbb{N}$).

Now, a two stage process that can be used for generation of Sophie Germain primes is going to be presented. In the first stage prime numbers are going to be produced by removal of all composite numbers from the set of natural numbers. In the second stage, we are going to analyze the prime numbers themselves, as a potential generators of odd primes. In the second stage all prime numbers that create composite numbers are going to be removed. Basically, we are going to implement two stage recursive process. At the end, only the prime numbers in the mps form, that represent the Sophie Germain primes, are going to stay. It is going to be shown that their number is infinite. It is easy to check that all numbers in mpl form are going to be removed from the set, based on the analysis made at the beginning of this chapter.

STAGE 1

Prime numbers can be obtained in the following way:

First, we remove all even numbers (except 2) from the set of natural numbers. Then, it is necessary to remove the composite odd numbers from the rest of the numbers. In order to do that, the formula for the composite odd numbers is going to be analyzed. It is well known that odd numbers bigger than 1, here denoted by a , can be represented by the following formula

$$a = 2n + 1,$$

where $n \in N$. It is not difficult to prove that all composite odd numbers a_c can be represented by the following formula

$$a_c = 2(2ij + i + j) + 1 = 2((2j + 1)i + j) + 1. \quad (1)$$

where $i, j \in N$. It is simple to conclude that all composite numbers could be represented by product $(i + 1)(j + 1)$, where $i, j \in N$. If it is checked how that formula looks like for the odd numbers, after simple calculation, equation (1) is obtained. This calculation is presented here. The form $2m + 1$, $m \in N$ will represent odd numbers that are composite. Then the following equation holds

$$2m + 1 = (2i + 1)(2j + 1) \quad ,$$

where $i, j \in N$. Now, it is easy to see that m must be in the form

$$m = 2ij + i + j. \quad (2)$$

When all numbers represented by m are removed from the set of odd natural numbers bigger than 1, only the numbers that represent odd prime numbers are going to stay. In other words, only odd numbers that cannot be represented by (1) will stay. This process is equivalent to the sieve of Sundaram [7].

The numbers that are left after this stage are prime numbers. If we denote with $\pi(n)$ number of primes smaller than some natural number n , the following equation holds [8]

$$\pi(n) \approx \frac{n}{\ln(n)}. \quad (3)$$

STAGE 2

Now, we should analyze numbers a that are left in observational space, or prime numbers themselves. With the exception of number 2 all other prime numbers are odd numbers. Since number 2 is Sophie Germain prime it will not be removed from the set. We are interested in removal of all numbers a that will create composite number when we generate number $2a + 1$. So, once more we are interested in removal of all numbers that generate composite odd numbers. So, once more we are going to implement (2) and remove all a in the form

$$a = 2ij + i + j. \quad (4)$$

That will leave us with prime numbers in mps form that represent the Sophie Germain primes. As it has been already explained, prime numbers in mpl form produce composite odd numbers divisible by 3, when formula $2 \cdot mpl + 1$ is applied on them, so they all are going to be removed in second stage. We denote sieve defined by (4) as SGP sieve, and number of Sophie Germain primes (SGP) with π_{SGP} .

Now, the other two stage sieve process is going to be presented. In the first stage of this process prime numbers are going to be produced. In the second stage the **indexes** of prime numbers are going to be processed. In the second stage only the prime numbers whose indexes are Mersenne numbers (numbers in the form $2^i - 1, i \in \mathbb{N}$) are going to stay. We are going to call those primes semi-Mersenne primes (or SMP). The sieve process in the second stage is going to be called Mersenne sieve. In the first step of Mersenne sieve all even indexes are going to be removed. Then, all odd indexes defined by the following equation are going to be removed ($p(i)$ is i -th prime number)

$$m_i = 2 p(i) j - 1, \quad i = 2, 3, 4, \dots, \quad (5)$$

where $j \in \mathbb{N}$. It is simple to understand that the number of Mersenne numbers (indexes) is infinite, since the number of Mersenne numbers smaller than some natural number n ($MN(n)$) is given by the following equation

$$MN(n) = \text{floor}(\log_2(n)) \quad (6)$$

Since it can be shown that the number of numbers left by the second stage sieve that produces Sophie Germain primes is bigger than the number of numbers left by Mersenne sieve, we can conclude that the number of Sophie Germain primes is infinite. In order to explain the previous conclusion we are going to compare sieves of the second stage of both processes, closely.

In the second stage, the only numbers left are prime numbers (actually, the stage one, as it is presented, is going to leave number 1 too, but it can be ignored). In the second stage of SGP generation, the sieve (4) is applied on indexes of odd natural numbers that are also primes. In the second stage of the SMP the Mersenne sieve is applied on the subset of natural numbers (natural numbers from 1 to the number of prime numbers) that represent indexes of prime numbers. In both cases, the same number of numbers (number of prime numbers) is analysed. In the following two tables the prime numbers and their indexes in both previously mentioned cases, are presented.

Table 1 Odd natural numbers (numbers in the form $2i + 1$) that are primes and their indexes that are available after the first stage of SGP generation

Index		2	3		5		7			...
Number	3	5	7		11	13		17	19	...

Table 1 - continuation

	11		13				17		19	...
	23			29	31			37		...

Table 2 Prime numbers and their indexes

Index	1	2	3	4	5	6	7	8	9	...
Number	2	3	5	7	11	13	17	19	23	...

Table 2 - continuation

10	11	12	13	14	15	16	17	18	19	...
29	31	37	41	43	47	53	59	61	67	...

We can now clearly see the difference between two processes. For instance, in the case of odd natural numbers that are prime, number 23 is generated by number 11, while in the case of prime numbers, number 23 is generated by index 9. However, in both cases number of numbers that are analysed and that are smaller or equal to 23 is 9. In the first case sieve is applied on the depleted set, while in the second case sieve is applied on the dense set.

Another significant difference is the number of threads that are necessary for the realization of the sieve in order to remove all unwanted numbers smaller than some natural number n . In the case of Mersenne sieve that number is equal to $\pi(\pi(n)/2)$ while in the case of second stage of SGP is $\pi(\sqrt{2n})$ (having in mind that realization of twin prime sieve in the second stage requires $\pi(\sqrt{n})$ sieves, it gives us a hint why there is always smaller number of SGP than the number of twin prime pairs smaller than some finite number n (starting from some number that is big enough), and why those values are asymptotically equal). It is not difficult to be shown that the following equation holds

$$\pi\left(\frac{\pi(n)}{2}\right) > \pi(\sqrt{2n}) \quad , \text{ for some } n \text{ that is big enough} . \quad (7)$$

The value of n in (7) that is big enough can be easily found by using the fact that prime counting function is non-decreasing and that square root function is strictly increasing (see Fig. 1).

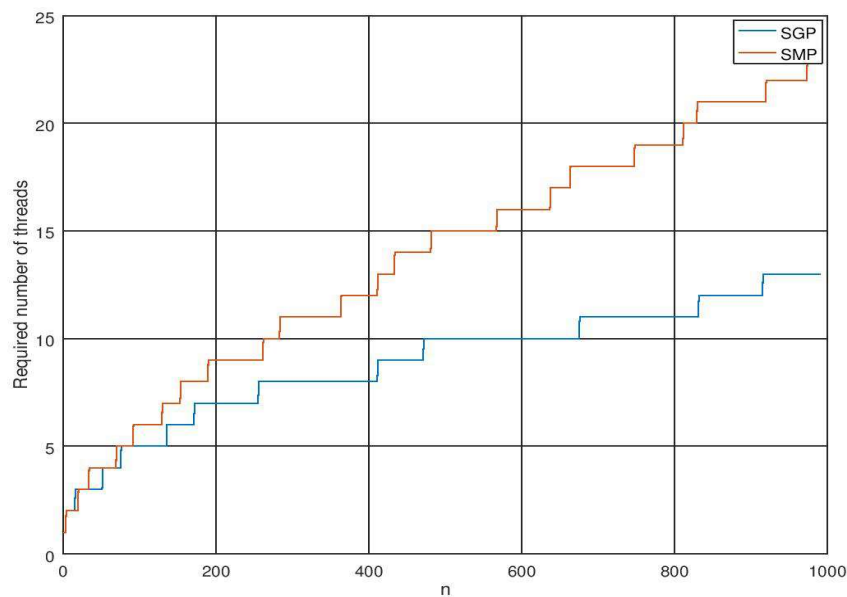


Fig. 1. Comparison of the number of threads required for the realization of the SMP and SGP sieves for generation of SMP and SGP numbers smaller than some natural number n

Here, the precise value of n is not of interest since we are not going to analyse cases in which n is finite number.

Now, we are going to analyze (4) and (5) in more detail. It can be seen that a in (4) and m in (5) are represented by the threads that are defined by odd prime numbers. For details see Appendix A. We are going to compare stages 2 step by step, for a few initial steps (analysis can be easily extended to any number of steps). The sieve in the second stage of generation of SGP is marked as SGPS.

Table 3. Comparison of the stages 2 for the generation of SGP and SMP – for threads defined by a few smallest primes

Step	MeS	Step	SGPS
1	Remove even numbers amount of numbers left 1/2	1	Remove numbers defined by thread defined by 3 (obtained for $i = 1$) amount of numbers left 1/2
	Thread: $m = 2j$		Thread: $a = 3j+1$
2	Remove numbers defined by thread defined by 3 (obtained for $i = 1$) amount of numbers left 2/3 of the numbers left in previous step	2	Remove numbers defined by thread defined by 5 (obtained for $i = 2$) amount of numbers left 3/4 of the numbers left in previous step
	Thread: $m = 3j - 1$		Thread: $a = 5j + 2$
3	Remove numbers defined by thread defined by 5 (obtained for $i = 2$) amount of numbers left 4/5 of the numbers left in previous step	3	Remove numbers defined by thread defined by 7 (obtained for $i = 3$) amount of numbers left 5/6 of the numbers left in previous step
	Thread: $m = 5j - 1$		Thread: $a = 7j + 3$
4	Remove numbers defined by thread defined by 7 (obtained for $i = 3$) amount of numbers left 6/7 of the numbers left in previous step	4	Remove numbers defined by thread defined by 11 (obtained for $i = 5$) amount of numbers left 9/10 of the numbers left in previous step
	Thread: $m = 7j - 1$		Thread: $a = 11j + 5$
5	Remove numbers defined by thread defined by 11 (obtained for $i = 5$) amount of numbers left 10/11 of the numbers left in previous step	5	Remove numbers defined by thread defined by 13 (obtained for $i = 6$) amount of numbers left 11/12 of the numbers left in previous step
	Thread: $m = 11j - 1$		Thread: $m = 13j + 6$

Note: In the previous table j represents natural numbers for thread $2j$, while for the other threads j represents odd or even numbers, so that corresponding thread defines odd numbers (all even numbers are removed by step one in first column and in the first stage for the second column).

Here, it has to be said that the values of the fractions in the Table 3 are asymptotically correct, but in the finite case they are only approximately correct. In the analysis that follows only the overall number of SGP is going to be analysed, since in that case we know that asymptomatic values are achieved.

From the Table 3 it can be noticed that threads defined by the same number in the first and second column will not remove the same percentage of numbers. The reason is obvious – consider for instance the thread defined by 3: in the first column it will remove $1/3$ of the numbers left, but in the second column it will remove $1/2$ of the numbers left, since the thread defined by 3 in stage 1 has already removed one third of the numbers (odd numbers divisible by 3 in observation space). So, only odd numbers (in observational space) that give residual 1 and -1 when they are divided by 3 are left, and there is approximately same number of numbers that give residual -1 and numbers that give residual 1, when the prime number is divided by 3. Same way of reasoning can be applied for all other threads defined by the same primes in different columns. More rigorous proof for the values of the fractions in second column (and for all other threads of SMPS that are not presented in the Table 3), based on Dirichlet's theorem on arithmetic progressions [9], is presented in Appendix B.

From Table 3 can be seen that in every step, except step 1, threads in the second column will leave bigger percentage of numbers than the corresponding threads in the first column. This could be easily understood from the analysis that follows:

- suppose that we have two natural numbers j, k such that $j - 1 \geq k$ ($j, k \in \mathbb{N}$), then the following set of equations is trivially true

$$j + k - 1 \geq 2k$$

$$-j - k + 1 \leq -2k$$

$$jk - j - k + 1 \leq jk - 2k$$

$$(j-1)(k-1) \leq (j-2)k$$

$$\frac{k-1}{k} \leq \frac{j-2}{j-1}$$

The equality sign holds only in the case $j = k + 1$. In the set of prime numbers there is only one case when $j = k + 1$ and that is in the case of primes of 2 and 3. In all other cases $p(i) - p(i-1) > 1$, ($i > 1$, $i \in \mathbb{N}$, $p(i)$ is i -th prime number). So, in all cases $i > 2$

$$\frac{p(i-1)-1}{p(i-1)} < \frac{p(i)-2}{p(i)-1} .$$

From Table 3 (or last equation) we can see that bigger number of numbers is left in every step of column 2 then in the column 1 (except 1st step). From that, we can conclude that after every step bigger than 1, number of the numbers that is left by the sieve defined by column 2 is bigger than number of numbers left by the sieve defined by the column 1 (that is also noticeable if we consider amount of numbers left after removal of all numbers generated by threads that are defined by all prime numbers smaller than some natural number).

From previous analysis, it is not difficult to understand that the following equation holds (p_{SMP} represents the number of semi-Mersenne primes, while $p_{SMP}(n)$ represents number of semi-Mersenne primes smaller than some natural number n – that means that indexes have to be smaller than $\pi(n)$)

$$\pi_{SGP} > p_{SMP} = \lim_{n \rightarrow \infty} p_{SMP}(n) .$$

Having in mind (6), and since it is easy to show that the following equation holds

$$\lim_{n \rightarrow \infty} \log_2(\pi(n)) = \infty ,$$

then it is easy to understand that the following equation holds

$$p_{SMP} = \lim_{n \rightarrow \infty} p_{SMP}(n) = \infty.$$

Now, we can safely conclude that the number of Sophie Germain primes is infinite. That concludes the proof.

3. Estimation of the number of Sophie Germain primes

Here we will state the following conjecture: for n big enough, number of Sophie Germain primes is given by the following equation

$$\pi_{SGP}(n) \sim 2C_2 \frac{\pi(n)}{\ln(\pi(n))},$$

where C_2 is twin prime constant [10]. Why it is reasonable to make such conjecture is explained in Appendix C. If we mark the number of primes smaller than some natural number n with $\pi(n) \approx f(n)$, where function $f(n)$ gives good estimation of the number of primes smaller than n , than $\pi_{SGP}(n)$, for n big enough, is given by the following equation

$$\pi_{SGP}(n) \sim 2C_2 \cdot f(f(n)).$$

If particular case $f(n) = Li(n)$, the following equation holds

$$\pi_{SGP}(n) \sim 2C_2 \cdot \int_2^n \left(\frac{d\pi(x)}{\ln(\pi(x))} \right) = 2C_2 \cdot \int_2^n \left(\frac{dx}{\ln(x) \ln\left(\int_2^x \left(\frac{dt}{\ln(t)}\right)\right)} \right).$$

For small number n , starting from the following formula for the prime numbers smaller than some natural number n

$$\pi(n) \approx \frac{n}{\ln\left(\frac{n}{2}\right) - \frac{1}{3}},$$

a good estimation of Sophie Germain primes smaller than natural number n is given by the

following formula

$$\pi_{SGP}(n) \approx (1 + 0.0129 \log(n)) \frac{n}{\left(\ln\left(\frac{n}{2}\right) - \frac{1}{3}\right) \left(\ln\left(\frac{n}{2}\right) - \ln\left(\ln\left(\frac{n}{2} - \frac{1}{3}\right)\right) + \frac{1}{3}\right)}.$$

This equation gives good approximation of the number of Sophie Germain primes smaller than natural number n , at least for the values of $n \leq 10^{14}$ (estimation for values $n = 10^3$ and $n = 10^4$ is actually correct).

References

- [1] T. Agoh. On Sophie Germain primes, Tatra Mt. Math. Publ. 20(65) (2000), 65-73.
- [2] M. Agrawal, N. Kayal, N. Saxena. PRIMES is in P, Ann. Of Math. 160(2)(2004), 781-793
- [3] R.A.J. Matthews. Maximally periodic reciprocals, Bull. Inst. Math. Appl. 28 (1992), 147-148.
- [4] W.-S. Tap, S.I. Yeo, S.-H. Heng, M. Henricksen. Security analysis of GCM for communication, Security and Communication Networks 7(5) (2014), 854-864.
- [5] H.M. Edwards. Fermat's Last Theorem: A Genetic Introduction to Algebraic Number Theory, Springer, 2000.
- [6] V. Shoup. A Computational Introduction to Number Theory and Algebra, Cambridge University Press, 2009.
- [7] V. Ramaswami Aiyar. Sundaram's Sieve for prime numbers, The Mathematics Student, 2(2) (1934), 73.
- [8] J.B. Rosser, L. Schoenfeld. (1962) Approximate formulas for some functions of prime numbers. Illinois Journal of Mathematics, 6(1), pp. 64-94.
- [9] P. G. L. Dirichlet. (1837) Bewies des Satzes, dass jede unbrengezte arithmetische Progression, deren erstes Glied und Differenz ganze Zahlen ohne gemeinschaftlichen Factor sind, unendlich

viele Primzahlen enthält. Abhandlungen der Königlich Preussischen Akademie der Wissenschaften zu Berlin, 48, pp. 45-71.

[10] G.H. Hardy, J.E. Littlewood. (1923) Some Problems of 'Partitio Numerorum.' III. On the Expression of a Number as a Sum of Primes, Acta Math. 44, pp.1-70.

[11] L. Euler. (1737) *Variae observationes circa series infinitas*. Comentarum Academiae Scientiarum Imperialis Petropolitanae, 9, pp. 160-188.

[12] G.L. Mullen, C. Mummert. (2007) *Finite Fields and Applications*, American Mathematical Society.

[13] C. Waid. (1974) On Dirichlet's Theorem and Infinite Primes, Proc. of the American Mathematical Society, 44(1), pp. 9-11.

APPENDIX A.

Here it is going to be shown that m in (2) is represented by threads defined by odd prime numbers.

Now, the form of (2) for some values of i will be checked.

$$\underline{\text{Case } i = 1:} m = 3j + 1,$$

$$\underline{\text{Case } i = 2:} m = 5j + 2,$$

$$\underline{\text{Case } i = 3:} m = 7j + 3,$$

$$\underline{\text{Case } i = 4:} m = 9j + 4 = 3(3j + 1) + 1,$$

$$\underline{\text{Case } i = 5:} m = 11j + 5,$$

$$\underline{\text{Case } i = 6:} m = 13j + 6,$$

$$\underline{\text{Case } i = 7:} m = 15j + 7 = 5(3j + 1) + 2,$$

$$\underline{\text{Case } i = 8:} m = 17j + 8,$$

It can be seen that m is represented by the threads that are defined by odd prime numbers. From examples (cases $i = 4$, $i = 7$), it can be seen that if $(2i + 1)$ represent a composite number, m that is represented by thread defined by that number also has a representation by the thread defined by one of the prime factors of that composite number. That can be proved easily in the general case, by direct calculation, using representations similar to (2). Here, that is going to be analyzed. Assume that $2i + 1$ is a composite number, the following holds

$$2i + 1 = (2l + 1)(2s + 1)$$

where $(l, s \in \mathbb{N})$. That leads to

$$i = 2ls + l + s.$$

The simple calculation leads to

$$m = (2l + 1)(2s + 1)j + 2ls + l + s = (2l + 1)(2s + 1)j + s(2l + 1) + l$$

or

$$m = (2l+1)((2s+1)j + s) + l$$

which means

$$m = (2l + 1)f + l,$$

and that represents the already existing form of the representation of m for the factor $(2l + 1)$, where

$$f = (2s + 1)j + s.$$

In the same way this can be proved for (4) and (5) (in that case represented by (5) analysis is even simpler).

APPENDIX B.

Now we are going to show that the inputs in the second column of Table 3 are correct. In order to do that, we are going to use Dirichlet's theorem on arithmetic progressions [9]. The theorem states that for any two positive coprime integers a and d , there are infinitely many prime numbers in the form $a + nd$, where n is also positive integer. Beside that, theorem also proves that for a given value of d , proportion of primes in each of progressions $a + nd$, asymptotically, is $1/\varphi(d)$, where $\varphi(d)$ represents Euler's totient function [11] that represents number of feasible progression for a given d , such that a and d are coprimes.

In the analysis that follows k represent natural number, while n represents nonnegative integer numbers. In order to simplify analysis, it is assumed that reader is capable to understand when certain context requires use of only odd or only even numbers n and/or k .

It is easy to understand that any thread in generative space, defined by some prime number in (4), will generate the thread in observational space that is defined by the same prime number, but with a different residual class. For instance, thread $3k$ in generative space will produce the thread $3(2k) + 1$ in observational space and so on. So, from now on, we are going to analyze numbers in observational space in order to make analysis easier.

Now, in the Step 1 of Stage 2 for generation of Sophie Germain primes, the numbers that are going to be removed are generated by a thread that is defined by prime number 3. That corresponds to the thread $3(2k)+1$ in observational space. In that case one half of the prime numbers are going to be removed. That follows directly from the Dirichlet's theorem [9], since all prime numbers can be expressed only in the form $3n+1$ or $3n+2$.

In the next step we are going to analyze what is going to happen when we remove thread defined by number 5, and which is given by $5(2k) + 3$ in observational space. In order to understand that, we are going to represent all numbers by the 15 threads defined by number 15. Those threads are

defined by the following progressions:

$15n+1$, $15n+2$, $15n+3$, $15n+4$, $15n+5$, $15n+6$, $15n+7$, $15n+8$, $15n+9$, $15n+10$, $15n+11$, $15n+12$,
 $15n+13$, $15n+14$ and $15n+15$.

We know that in the first stage (generation of prime numbers) numbers divisible by 3 ($1/3$ of the threads) defined by threads $15n+3$, $15n+6$, $15n+9$, $15n+12$ and $15n+15$ are going to be removed, as well as numbers divisible by 5 ($1/5$ of the threads left) defined by the threads $15n+5$ and $15n+10$.

The threads that are left are:

$15n+1$, $15n+2$, $15n+4$, $15n+7$, $15n+8$, $15n+11$, $15n+13$ and $15n+14$.

Based on Dirichlet's theorem we know that each of these threads contain $1/8$ of the prime numbers.

In the first step of second stage odd numbers in the form $3k + 1$, should be removed. That means that half of the threads that can be represented in the form $3k + 1$ are going to be removed. Those threads are $15n+1$, $15n+4$, $15n+7$ and $15n+13$.

So, the threads that are left are $15n+2$, $15n+8$, $15n+11$ and $15n+14$, and each of them contain $1/8$ of the prime numbers.

In the second step of the second stage, numbers defined by thread defined by 5, in the form $5(2k)+3$ have to be removed. The only thread that was left and that can be expressed in the form $5(2k)+3$ is thread $15n+8$. Now, we can easily conclude that number of primes that is removed by the thread defined by number 5 is $1/4$ of the numbers left. That means that $3/4$ of the numbers will be left.

However, it is difficult to generalize the proposed method for the other steps in Stage 2. So, an alternative method is going to be analyzed.

After all numbers in the form $3(2k) + 1$ are removed, we know that all odd prime numbers that are left have to be in the form $3(2k+1) + 2$, or, for the sake of simplicity, in the form $3k + 2$ (and reader should have in mind that we are talking only about odd numbers, since all even numbers were removed in the first step of Stage 1). We know that all numbers that are left have to be in some of

the following forms

$3(5n+1) + 2, 3(5n+2) + 2, 3(5n+3) + 2, 3(5n+4) + 2$ and $3(5n+5) + 2$, or

$$15n + 3*1 + 2, 15n + 3*2 + 2, 15n + 3*3 + 2, 15n + 3*4 + 2, 15n + 3*5 + 3*0 + 2. \quad (B.1)$$

Since all forms in (B.1) contain the term(s) divisible by 15 (and consequently divisible by 5), it is clear that additional forms that are going to be removed, will be removed based on the analysis of the following expressions

$$3*1 + 2, 3*2 + 2, 3*3 + 2, 3*4 + 2, 3*0 + 2. \quad (B.2)$$

We know that in the first stage thread that is divisible by 5 has to be removed and in the second step of the Stage 2, thread that is in the form $5k + 3$, has to be removed. We can see that all five terms in equation (B.2) represent simple calculations on the finite field Z_5 [12]. It is known that in that case, multiplication of all elements of the field with element of the field that is not zero, will lead to a permutation of the elements of the field [12]. Also, addition of the one nonzero element of the field to all other elements of the field will lead to a permutation of the elements of the field [12]. From that we can conclude that exactly one term will be congruent to 0 by modulo 5, and only one term will be congruent to 3 by modulo 5. That means that out of 5 threads defined by (B.1), three are going to stay after second step in Stage 2, which means that $\frac{3}{4}$ of the numbers that were left after step 1 in Stage 2, are going to stay after removal of the corresponding thread defined by number 5 (that is based on the Dirichelt's theorem [9] - all feasible threads defined by number 15 contain the same number of prime numbers).

After step 2 in Stage 2, all numbers can be written in the following forms

$$15n + 2, 15n + 11 \text{ and } 15n + 14. \quad (B.3)$$

The proposed analysis can be applied on all consecutive step of Stage 2. Now, in the step 3 of the Stage 2, we are going to apply a similar analysis like in the step 2 of Stage 2. In this case, instead of one thread defined by $3k + 2$, we have three threads defined by (B.3). In the third step of Stage 2, thread defined by number 7 is going to be removed. Impact of that removal is the easiest if we

analyze the following forms of the remaining threads (here we are going to present forms for thread $15k + 2$; the other 2 threads could be analyzed analogously)

$$15(7n+1)+2, 15(7n+2)+2, 15(7n+3)+2, 15(7n+4)+2, 15(7n+5)+2, 15(7n+6)+2, 15(7n+7)+2, \text{ or}$$

$$105n+15*1+2, 105n+15*2+2, 105n+15*3+2, 105n+15*4+2, 105n+15*5+2, 105n+15*6+2, 105n+15*7+15*0+2. \quad (\text{B.4})$$

Since all forms in (B.4) contain the term(s) divisible by 105 (and consequently divisible by 7), it is clear that additional forms that are going to be removed, will be removed by analysis based on the following expressions

$$15*1+2, 15*2+2, 15*3+2, 15*4+2, 15*5+2, 15*6+2, 15*0+2, \quad (\text{B.5})$$

or having in mind that $a*b \pmod{7} = a \pmod{7} * b \pmod{7}$, the forms of interest are given by the following equation

$$1*1+2, 1*2+2, 1*3+2, 1*4+2, 1*5+2, 1*6+2, 1*0+2. \quad (\text{B.6})$$

Similarly to the situation in step 2, we can see that all seven terms in equation (B.6) represent simple calculations on the finite field Z_7 [12]. Using the same line of reasoning like in the previous step, we can conclude that fraction of number of numbers that are going to stay after step 3 is exactly the one given in Table 3, and that is $5/6$ of all numbers left after step 2 (here is assumed that the same analysis can be analogously performed for the other 2 threads defined by (B.3)). After this step 15 threads defined by number 105 are going to stay and each is going to contain the same percentage of prime numbers.

Now, it is obvious that proposed analysis can be applied to all consecutive steps of stage 2. In all cases, the removal of certain threads will be based on multiplication and addition of the finite field Z_{pk} , where pk represents the odd prime number that defines thread that is going to be removed in the k -th step of Stage 2. In all cases those multiplications and addition will result in the permutation

of all elements of the corresponding finite field and it can be shown that in every step they are going to leave the ratio $(pk-2)/(pk-1)$ of available numbers, by using reasoning similar to the cases $pk = \{3, 5, 7\}$. From this analysis we can understand that the values presented in the second column of Table 3 are correct. Same can be concluded for all other threads that are not presented in the table. The proposed analysis holds also in the case of threads that are defined by prime numbers that are infinite (see [13]).

APPENDIX C.

Here, asymptotic density of numbers left, after implementation of the eSuS and the SGP sieve is going to be calculated (the eSuS represents extended SuS⁰ sieve, in which after the removal of thread defined by some prime number in (2), also that prime number is removed). After removal of all even composite numbers and first k steps of the SuS⁰ sieve, density of numbers left is given by the following equation

$$c_k = \frac{1}{2} \prod_{j=2}^{k+1} \left(1 - \frac{1}{p(j)}\right),$$

where $p(j)$ is j -th prime number. (We should have in mind that the density considered in previous equation are asymptotically correct).

In the case of SGP sieve the density of numbers left after the first k -steps is given by the following equation

$$c2_k = \prod_{j=2}^{k+1} \left(1 - \frac{1}{p(j)-1}\right) = \prod_{j=2}^{k+1} \left(\frac{p(j)-2}{p(j)-1}\right).$$

So, if implementation of first sieve will result in the number of prime numbers smaller than n which we denote as $\pi(n)$, than implementation of the second sieve on some set of size $\pi(n)$ should result in the number of numbers $gp(n)$ that are defined by the following equation (for some big enough n)

$$gp(n) = r_{S2SI}(n) \cdot \frac{\pi(n)}{\ln(\pi(n))},$$

where $r_{S2SI}(n)$ is defined by the following equation (k is the number of primes smaller or equal to $\sqrt{2n}$)

$$r_{S2SI}(n) = \frac{c2_k}{c_k} = \frac{\prod_{2 < p \leq \sqrt{2n}} \left(\frac{p-2}{p-1}\right)}{\prod_{p \leq \sqrt{2n}} \left(\frac{p-1}{p}\right)} = 2 \prod_{2 < p \leq \sqrt{2n}} \left(\frac{p-2}{p-1}\right) \left(\frac{p}{p-1}\right) \approx 2C_2.$$

where p represents prime number. For n that is not big, $gp(n)$ should be defined as

$$gp(n) = f_{COR}(n) \cdot 2C_2 \cdot \frac{\pi(n)}{\ln(\pi(n))},$$

where $f_{COR}(n)$ represents correction factor that asymptotically tends toward 1 when n tends to infinity. Here, the number of numbers left after eSuS is approximated by $n/\ln(n)$ (see equation (3)).