Finding Clusters of Primes, I Progress Report 2003 - 2005

Jörg Waldvogel and Peter Leikauf Seminar for Applied Mathematics SAM Swiss Federal Institute of Technology ETH, CH-8092 Zürich

January 2003

1 Purpose

In the current application of our parallelization concept we are using an algorithm involving sieving techniques for locating and counting clusters of prime numbers. Whereas the distribution of primes seems to be fairly regular (if the Riemann hypothesis is true), the distribution of twin primes and longer clusters is largely unknown and is characterized by large-scale anomalies. Collecting experimental data on these anomalies is one of the reasons for the interest in clusters of primes.

Another challenge of finding clusters of primes is collecting data supporting the unproven *prime k-tuple hypothesis*, which is concerned with sequences

$$\mathbf{c} = [c_1, c_2, \dots, c_k], \quad c_i \in \mathbb{Z}, \quad j = 1, 2, \dots, k, \quad c_1 < c_2 < \dots < c_k$$

of k integers. The sequence or pattern \mathbf{c} is called admissible if its elements leave at least one residue class empty modulo every prime. As a consequence of this definition, simple divisibility considerations cannot exclude the existence of integers x such that each element of the shifted pattern $\mathbf{c}_x = [x + c_1, x + c_2, \dots, x + c_k]$ is prime. A shifted pattern \mathbf{c}_x with this property will be referred to as a prime instance of \mathbf{c} . The prime k-tuple hypothesis states that every admissible pattern has infinitely many prime instances.

An example of a non-admissible pattern is $\mathbf{c} = [0, 2, 4]$; modulo 3 all three residue classes are occupied. For every integer x the shifted pattern $\mathbf{c}_{\mathbf{x}} = [x, x+2, x+4]$ contains one multiple of 3; therefore there exist only finitely many prime instances, namely for x=3 and x=-7, when one of the elements is the prime 3 itself or its opposite.

The goal of this project is to find and display prime instances of certain admissible patterns, in particular of the densest possible patterns for a given number n of elements. With the current software and hardware $n \leq 21$ and $|x| \leq 10^{30}$ delimit the region of feasible search.

Many dense admissible patterns may be defined by a finite subsequence of consecutive positive or negative odd primes. E.g. $\mathbf{c} = [17, 19, 23, 29, 31]$ is admissible, it reoccurs as the prime instance $\mathbf{c}_{210} = [227, 229, 233, 239, 241]$, and many more may be found. In general, the inverse (i.e.the mirror image) of an admissible pattern is again admissible: $\mathbf{c}' = [-31, -29, -23, -19, -17]$ has the prime instance $\mathbf{c}'_{1320} = [1289, 1291, 1297, 1301, 1303]$, etc.

In this text, admissible patterns will be denoted by listing the elements c_j in increasing order, separated by commas and enclosed in square brackets (often normalized such that $c_1 = 0$). For brevity we will abbreviate a list of consecutive primes by merely indicating the bounding elements, seprated by two periods: $\mathbf{c} = [17..31]$, $\mathbf{c}' = [-31..-17]$.

A proof of the prime k-tuple hypothesis is currently out of reach; not even for the simplest case, the twin prime hypothesis, a proof is in sight. In contrast, the observed average densities of prime k-tuples in the accessible range are in perfect agreement with the densities $\rho_{\mathbf{c}}(x)$ conjectured by Hardy and Littlewood [5] in 1922:

(1)
$$\rho_{\mathbf{c}}(x) = \frac{h_{\mathbf{c}}}{(\log x)^{|\mathbf{c}|}},$$

where $h_{\mathbf{c}}$ is the Hardy-Littlewood constant associated with the pattern \mathbf{c} , and $|\mathbf{c}|$ is the number of elements in the pattern \mathbf{c} .

2 Review of Results 2001 - 2003

The main achievement in the preceding period was the discovery of two dense clusters of 18 primes in the range of $3 \cdot 10^{24}$ on November 13, 2000 and on January 31, 2001. This computation was about 50 times harder than finding the 23-digit clusters of 17 primes among 67 consecutive integers

(first discovered by Tony Forbes [3], and indepently by J. Waldvogel [10], both in 1998). The news about this discovery was immediately announced in the number theory press [6]. It also got coverage in the web journal of ETH (ETH Life, December 6, 2000, [2]). Dense clusters of primes receive particular attention on the website [4], continuously actualized by Tony Forbes. The successes of our implementation bear the danger of monopolizing this site and taking away all the fun!

We approched our "prime" target, the search for clusters of 18 primes among 71 consecutive integers, by searching blocks of size 10^{24} . Preliminary experiments with the idle time of 20 workstations of the Seminar for Applied Mathematics SAM were carried out for testing purposes. The time necessary in this setting was estimated as 2 years – with luck; otherwise it could be 5 years as well. If the search were to be successful within a reasonable timespan, we could claim a world record that wouldn't be easy to break. Dense clusters of 19 primes in the patterns of $[13, \ldots, 89]$, $[37, \ldots, 113]$ or their mirror images are expected to repeat/occur only in the range of 10^{25} or even 10^{26} . For the final "attack" on the 18er 432 processors of the Beowulf Cluster and 20 workstations of SAM were involved. The search has now been completed up to $2.9999949836 \cdot 10^{24}$. The first search was caried out with the pattern

(2)
$$\mathbf{c} = [x - 83, \ x - 79, \ x - 73, \ \dots, \ x - 19, \ x - 17, \ x - 13]$$

consisting of 18 elements, such that for x=0 the sequence of consecutive negative primes beginning at -83 and ending at -13 is obtained. For the corresponding Hardy-Littlewood constant the value $h_c=6723654.312$ is obtained. Equ. (1) implies that the expected number of occurrences of the pattern \mathbf{c} (as well as of its mirror image) in a large interval of length Δ near a much larger x is approximately given by $h_c \Delta (\log x)^{-18}$. Integration of this average density yields the following expected frequencies $\mathrm{HL}(x)$ (referred to as the Hardy-Littlewood count, HL count for short):

$$x$$
 $1_{10}24$ $2_{10}24$ $3_{10}24$ $4_{10}24$ $5_{10}24$ $1_{10}25$ $1_{10}26$ $HL(x)$ 0.438 0.695 0.912 1.107 1.286 2.056 9.962

Therefore, one could "hope" that the above prime pattern \mathbf{c} would repeat for some x in the range of $3 \cdot 10^{24}$, and also that its mirror image occurs in the same range. One could even be lucky to have occurrences for smaller values of \mathbf{x} , but also, with bad luck, the search might have to be pushed up

much higher. The table below summarizes some of the actions that lead to the discovery of the two 18-tuples of maximum density; for more details see [10], Projects/cl18.pdf.

Pattern	Block	Begin/end of	search Date	Initial element	HLcount
[-8313]] 1	8/03 - 9/19/	′00		0.438
[-8313]] 2	9/19 -10/26/	′00		0.695
[-8313]] 3	10/29 -11/20/	00 11/13/00	284537254250991186826680	0.880
[13 83]] 1	12/19 - 1/23/	' 01		0.438
[13 83]] 2	1/23 - 2/27/	01 1/31/01	190623083504664829329004	13 0.673
[13 83]] 3	2/27 - 3/26/	'01		0.912

3 Distribution of a Particular 12-Tuple

In this section we describe the results of an extended study of the distribution of the particular k-tuple with k = 12, defined by the pattern

(3)
$$\mathbf{c} = [-23, -19, -17, -13, -11, -1, 1, 11, 13, 17, 19, 23].$$

It is easily seen that \mathbf{c} is admissible: modulo the primes ≤ 7 the residue class 0 is vacant, modulo the primes ≥ 13 the 12-tuple necessarily leaves at least one residue class empty, and modulo 11 the residue classes ± 4 are empty. The pattern \mathbf{c} , spanning an interval of length 46, is not the densest admissible 12-tuple; the primes $[11,13,\ldots,53]$, spanning an interval of length 42 only, also form an admissible 12-tuple. The pattern \mathbf{c} , by its appealing symmetry and by the lack of an obvious reoccurrence in terms positive primes, promises to be an interesting object to study.

The prime k-tuple hypothesis predicts infinitely many prime instances of \mathbf{c} . By means of our algorithm ([10]: the Initial Report 2001, /clprimes01.pdf, and the experimental PARI code, /paricode.gp) all prime intances of \mathbf{c} up to $1.000046276 \cdot 10^{21}$ were generated, 155 898 instances alltogether. Chinese remaindering was done with the first $\mathbf{np} = 14$ primes. The primes up to 29 produce 52 224 tasks; the range mentioned above was covered with 5 sieve blocks of length $\mathbf{block} = 15288$. These computations were performed during the testing phase of the algorithm on Asgard within about two weeks.

In the table below we report the initial primes of a few of the patterns generated: 16 at the beginning of the sequence, 7 at the current end, and 2 at each transition to a new decade.

Index	Initial Prime	Index	Initial Prime
1	41280160361347	205	99859425208272637
2	65073487398967	206	100306705532373247
3	273596722858597		
4	305247832189207	996	998054166528529927
5	314546191059007	997	1004612024526939967
6	334701417639727		
7	355340244780337	5330	9999309335149183957
8	552105775370287	5331	10001687719716943297
9	610954727181937		
10	660119078678197	28104	99992682101800188517
11	767920116273217	28105	100004729076950887327
12	776137769447857		
13	899316460923697	155892	999972884065149535357
14	957065931972967	155893	1000006962511995788257
15	1333086396411817	155894	1000017025199672760187
16	1490344945469287	155895	1000022902668956065807
		155896	1000035195619662051757
43	9411514972889167	155897	1000037569658302164127
44 1	0012543630829647	155898	1000044549710737801177

In the following, we consider the counting function $\pi_{\mathbf{c}}(x)$, defined as the number of patterns \mathbf{c} with initial prime $\leq x$. We will compare $\pi_{\mathbf{c}}(x)$ with the corresponding smooth counting function obtained by integrating the Hardy-Littlewood density given in Equ. (1). For the pattern \mathbf{c} of Equ. (3) the Hardy-Littlewood constant

$$h_{\mathbf{c}} = 18867.657182534569614833826514092$$

is obtained. Instead of the conventional logarithmic integral

$$\operatorname{li}_k(x) := \int_2^x \frac{1}{(\log t)^k} \, dt$$

we prefer to use the generalization

(4)
$$R_k(x) := \sum_{j=0}^{\infty} \frac{(\log x)^j}{(k-1+j)! \, j \, \zeta(1+j)}, \quad k \in \mathbb{N},$$

of Riemann's function $R(x) := R_1(x)$. Here $\zeta(s)$ is the Riemann zeta function, and for j = 0 the product $j \zeta(1+j)$ must be understood in the sense of the limit $\lim_{j\to 0} (j \zeta(1+j)) = 1$.

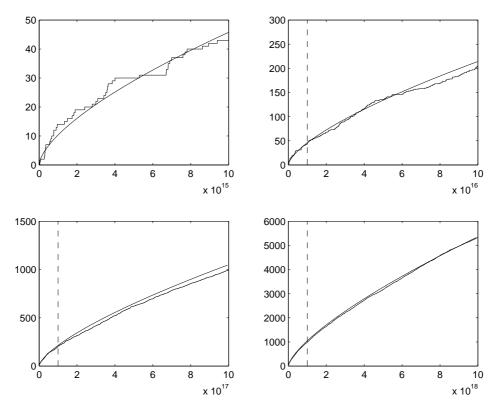


Figure 1: Counting function $\pi_{\mathbf{c}}$ and the smooth graph of $\mathrm{HL}(x)$ versus x. Range $0 < x < 10^{j}, \ j = 16, \ldots, 19$.

For $x \to \infty$ $R_k(x)$ and $\mathrm{li}_k(x)$ have asymptotically the same values. The advantage of $R_k(x)$ is the simple evaluation algorithm based on the series (4) and the regularity at x = 1. $R(x) = R_1(x)$ is known to be a very elegant monotonic approximation of the prime counting function $\pi(x)$. In all our examples the (modified Hardy-Littlewood) counting function

(5)
$$\operatorname{HL}(x) := h_{\mathbf{c}} \cdot R_k(x) \quad \text{with} \quad k = |\mathbf{c}|$$

agrees surprisingly well with $\pi_{\mathbf{c}}$, for large values of x as well as for small ones.

In the six frames of Fig. 1, Fig. 2, Fig. 3 we plot $\pi_{\mathbf{c}}$ and $\mathrm{HL}(x)$ in the regions $0 < x < 10^{j}, \ j = 16, \ldots, 21$. The range of x in the subsequent frame is increased 10-fold. The preceding frame appears to the left of the dashed lines in compressed form.

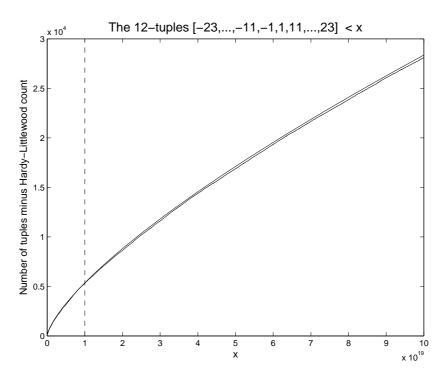


Figure 2: Counting function π_c and the smooth graph of HL(x) versus x. Range $0 < x < 10^{20}$. π_c below HL.

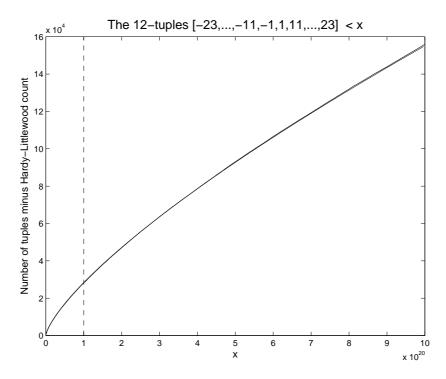


Figure 3: Counting function $\pi_{\bf c}$ and the smooth graph of ${\rm HL}(x)$ versus x. Range $0 < x < 10^{21}$. Crossover of $\pi_{\bf c}$.

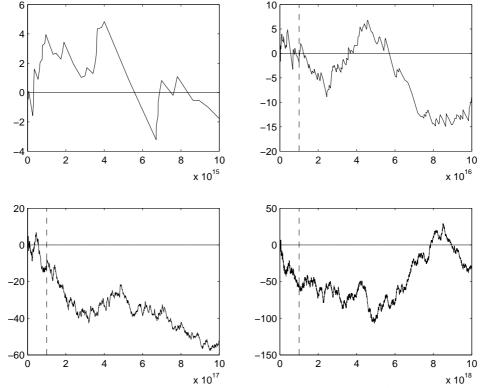


Figure 4: Difference $\pi_{\mathbf{c}} - \mathrm{HL}(x)$ versus x. Range $0 < x < 10^{j}, \ j = 16, \dots, 19$.

In the six frames of Fig. 4, Fig. 4, Fig. 6 we plot the difference $\pi_{\mathbf{c}} - \mathrm{HL}(x)$ versus x in the same regions $0 < x < 10^{j}$, $j = 16, \ldots, 21$. Again, the range of x in the subsequent frame is increased 10-fold. The preceding frame appears to the left of the dashed lines in compressed form.

Despite apparent discrepancies between $\pi_{\mathbf{c}}$ and $\mathrm{HL}(x)$ (e.g. in the range 10^{20} , Fig. 5) the agreement of the two graphs over the full range of 7 powers of 10 is striking (Fig. 3). Whereas the prime counting function $\pi(x)$ oscillates on a short scale, the counting functions of prime twins and longer clusters of primes seem to depart from the average in large-scale excursions. E.g., the pattern \mathbf{c} of Equ. (3) is deficient in $9 \cdot 10^{18} < x < 2.7 \cdot 10^{20}$, i.e. the bounds of this interval differ by as much as a factor of 30. Even in the simplest case, the prime twins, the source of such large-scale anomalies is still unknown and is the subject of theoretical research.

In Fig. 7, the final graphic of this section, we summarize the properties of $\pi_{\mathbf{c}}$ by using a logarithmic scale in x and plotting the normalized difference

$$D(x) = \frac{\pi_{\mathbf{c}} - \mathrm{HL}(x)}{\sqrt{\mathrm{HL}(x)}}.$$

In the range considered this choice of the normalizing factor seems to keep the amplitude roughly constant; however, its true growth rate is not known.

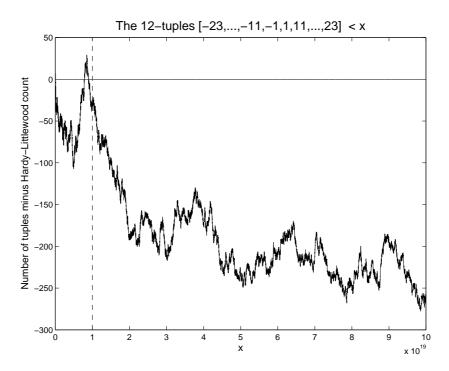


Figure 5: Difference $\pi_{\mathbf{c}} - \mathrm{HL}(x)$ versus x. Range $0 < x < 10^{20}$.

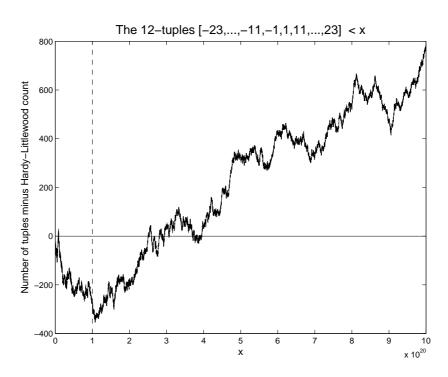


Figure 6: Difference $\pi_{\mathbf{c}} - \mathrm{HL}(x)$ versus x. Range $0 < x < 10^{21}$.

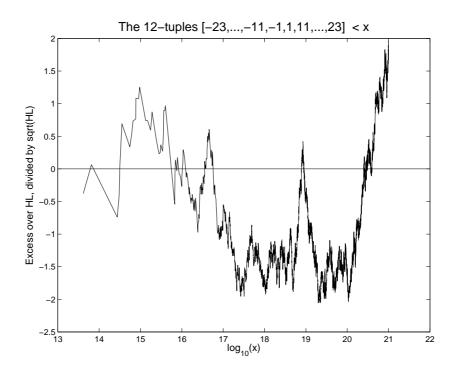


Figure 7: Normalized difference $(\pi_c - \text{HL}(x))/\sqrt{\text{HL}(x)}$ versus $\log_{10}(x)$. Range $10^{14} < x < 10^{21}$, logarithmic scale.

4 Some Particular Patterns

4.1 Densest 16-Tuples

The densest 16-tuples occur in two patterns which may be defined by sequences of consecutive small primes, namely

$$\mathbf{c} = [13, 17, \dots, 71, 73]$$
 and $\mathbf{c}' = [-73, -71, \dots, -17, -13]$,

their common Hardy-Littlewood constant being $h_c = 751221.42544528571$. Below we will compare the modified Hardy-Littlewood counting function

$$HL(x) = h_{\mathbf{c}} \cdot R_{16}(x)$$

with the actual counting functions $\pi_{\mathbf{c}}(x)$, $\pi_{\mathbf{c}'}(x)$ deduced from the 67 instances present in $0 < x < 10^{23}$.

	Pattern [13,,73]	Pattern [-73,,-13]
1	13	47710850533373130107
2	695874886175252911063	347709450746519734877
3	1567582627835236839763	1099638576123052218257
4	1750052554011927712483	1169914227530138703617
5	2257588388550898970503	1522014304823128379267
6	3789227751026345304613	1620784518619319025977
7	4654682384109074514133	2639154464612254121537
8	5022156579757255625623	3259125690557440336637
9	13599236099159166553033	9042634271485192050677
10	29894522822363684652103	9239395687646993061197
11	35718904544536715448883	15571053758048293307807
12	42421183685552747462323	20628149050698694668167
13	47624415490498763963983	20947353617877810296177
14	50069823850049036630533	23182160505954925788317
15	56294926786180569503953	27814116054901200587567
16	60877851518090858117803	30406149669349341460577
17	66871135379953148611303	31607383424682394081757
18	70743491366529526461853	34254730511961158822627
19	72039555441202354852783	40675973411840597813987
20	72939169778564978895943	41459159264655740911307
21	78314167738064529047713	49798002215047773229547
22	79415821643818505392483	53284658140441622257367
23	80755924819458605984203	61196813406933554172827
24	81433995543774820773763	62564565965531138812307
25	83405687980406998933663	67089635430601104554237
26	93091355499511979496823	73377493322819357084417
27	94045986419037158522953	75356498508757177334627
28	94975866735959660878363	80444001154123052328887
29	97907646478336552814893	81044576768686984666217
30	99517467925153764522973	82069568461982909560757
31		86042715202634832665027
32		88077044957120569823477
33		89172493451857275102647
34		91576165734484982223677
35		93954295353554241598997
36		97692369541610844803807
37		98659593721870389301097

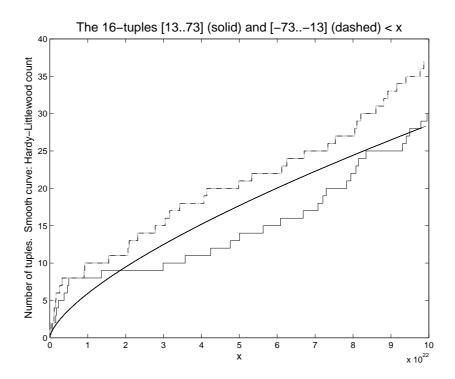


Figure 8: The densest 16-tuples in the range $0 < x < 10^{23}$ and their HL count. Solid: Pattern [13,...,73], dashed: Pattern [-73,...,-13].

4.2 Two Octuples at Minimum Distance

The densest octuples occur in three patterns which may all be defined by sequences of consecutive small primes, namely

$$\mathbf{c_1} = [17, \dots, 43], \ \mathbf{c_2} = [11, \dots, 37], \ \mathbf{c_2}' = [-37, \dots, -11].$$

 $\mathbf{c_1}$, spanning a range of 26, is the longest symmetric pattern of maximum density. We observe that the 16-tuple \mathbf{C} consisting of two copies of $\mathbf{c_1}$, concatenated at the distance of 34,

$$\mathbf{C} = [-43, -41, -37, -31, -29, -23, -19, -17, 17, 19, 23, 29, 31, 37, 41, 43],$$

is admissible (see the remark at the beginning of Section 3). For the corresponding Hardy-Littlewood constant $h_{\bf C}$ we obtain

$$h_{\mathbf{C}} = \frac{13312}{2625} \cdot h_{\mathbf{c}} = 3809622.71067719746,$$

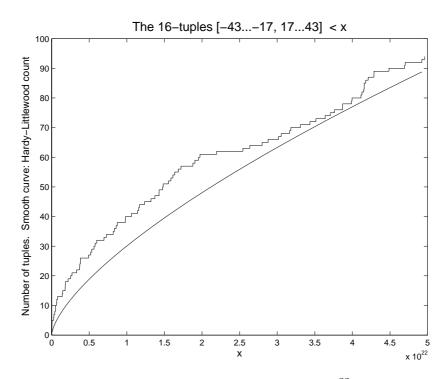


Figure 9: 94 pairs of octuples in the range $0 < x < 5 \cdot 10^{22}$ and their HL count, Pattern [-43,...,-13, 13,...,43].

where $h_{\mathbf{c}}$ is defined in Section 4.1. Up to the limit $X = 5 \cdot 10^{22}$ we expect $\mathrm{HL}(X) = h_{\mathbf{C}} \cdot R_{16}(X) = 89.7$ occurrences of the pattern \mathbf{C} . The distribution of the 94 prime instances of \mathbf{C} is shown in Fig. 9.

The patterns

$$\mathbf{C_1} = [-43, -41, -37, -31, -29, -23, -19, -17, -1, 1, 17, 19, 23, 29, 31, 37, 41, 43, 47] \\ \mathbf{C_2} = [-47, -43, -41, -37, -31, -29, -23, -19, -17, -13, 1, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47]$$

of 18 or 21 elements, respectively, both containing \mathbf{C} as a subset, are found to be admissible as well. Therefore, it is expected that some of the 94 instances of \mathbf{C} will actually be subsets of $\mathbf{C_1}$ or $\mathbf{C_2}$ with more than 16 elements. However, instances of $\mathbf{C_1}$ or $\mathbf{C_2}$ themselves are expected to occur only in the range of 10^{24} or 10^{30} , respectively. In the table below we list the initial primes of all 94 instances of \mathbf{C} , together with additional elements from $\mathbf{C_1}$ or $\mathbf{C_2}$ contained in the patterns, if present (in the columns more). No patterns with more than 18 elements exist in the range considered.

Tne	lor Initial Drive	momo	Tndov	Tnitial Dmima	mama
1 nc	lex Initial Prime 10458834002271815117	more	Index	Initial Prime 14284612658991181216667	more
2	26476006821087640697	-47		14297146316095286150387	
3	44350865905809142637	-41		14754208002759013618787	
		47 10			10
4	54014646858393564377	-47,13		14841953373654341817917	-13
5	62155369550078511587	-47		15532484963882491131407	-13
6	253586253591518370557	-47		15853064083490265705107	47
7	304079924911990894547	13		16130615338665630649727	
8	423291158347150012877			16372205147465011401437	
9	511505988322414165037			16858454028003145410887	
10	512761727903842750367			17198991227897242847117	
11	644424770171034352457			18786403726068543935327	4.0
12	675759858713748355427			19054062004470982522067	13
13	780362378270548056017			19547070837095799260297	
14	1416058157129915879537	457		19728243594481889687177	
15	1457922032650513232837	-47		21944528505231528596357	
16	1792491363005354644667			25435635163943029705637	1
17	1812456008867090693987			26360513714179113726137	
18	1826785107639242841047			28001420822911641278207	
19	2193719983731075106187			28794590387631746120477	
20	2519939592777795291137			30178895046554873882297	
21	2715485152912746884627			30544994855724415219187	
22	3309935724597716754497			31647354432491850512417	-1
23	3703048128556987693517			31810951085696093237807	
24	3741636047391669917447	-47,-1		33086235745772630878337	
25	3824778120476297871767			34351709846301758103227	
26	3829402050773411044967			35138861572319523491507	4.
27	4916234207084567704217			36373617998432758393307	-47
28	5187457951820816722007			37064354400225808655957	-47
29	5335829523905291634257			37604920347436702417037	13
30	5644872371007238806317			38694494289512089895747	-47
31	5795560513558228593137			38695164866270310508067	-47
32	5984674091995810261007			39874467007461049653647	
33	6969682106336094492227			39950366639130114006797	-47
34	7305028693989548271707			41132274627392276695247	
35	8203874622781421977427	-13		41269649010979497985757	
36	8348101522292016031427			41504939256100372845827	
37	8591703894674746255367	47		41564901882588422431127	-47
38	8741497840086683458667	1		41579404447063334354537	
39	9806894079711946604267	1		41773491218296251541007	
40	9815343014825749197677			42108300110425746630467	
	10606325734168483068707	-47		42858583357347352501787	
	11423347423634306069867			42876198440221191818807	
	11568680835710656040417			44860553600587843401317	
	11673556306524425252897	1		46952529647970814520687	
	12344874940525148664677			47030198166588987767297	
	13216411836106066110557			49235705631942915972767	
47	13729297230363654379397	47	94	49623872890924546023767	

References

- [1] C. Batut, K. Belabas, D. Bernardi, H. Cohen, M. Olivier: *The software package PARI* (freeware). http://pari.math.u-bordeaux.fr/
- [2] ETH Life, Die tägliche Webzeitung der ETH, 6. Dezember 2000, Archiv. *Prozessoren malochten 100 Tage*. http://www.ethlife.ethz.ch/articles/tages/Waldvogel.html
- [3] Tony Forbes: *Prime clusters and Cunningham chains*. Math. of Comp. **68** (1999) 1739-1747.
- [4] Tony Forbes: Prime k-tuplets. http://www.ltkz.demon.co.uk/ktuplets.htm
- [5] G.H. Hardy and J.E. Littlewood: Some problems of Partitio Numerorum III. Acta Math. 44, 1922, 1-70.
- [6] Number theory news. http://www.utm.edu/research/primes/
- [7] Paulo Ribenboim: The New Book of Prime Number Records, 3rd ed. Springer 1996, 541 pp.
- [8] Hans Riesel: Prime Numbers and Computer Methods for Factorization, 2nd ed. Birkhäuser 1994, 464 pp.
- [9] R.L. Rivest, A. Shamir, L. Adleman: A method for obtaining digital signatures and public-key cryptosystems. Communications of the ACM 21 (1978) 121-126.
- [10] Jörg Waldvogel: Homepage. http://www.math.ethz.ch/~waldvoge/Projects/