The Computation of Previously Inaccessible Digits of π^2 and Catalan's Constant

David H. Bailey* Jonathan M. Borwein† Andrew Mattingly‡
Glenn Wightwick§

June 10, 2013

1 Introduction

We recently concluded a very large mathematical calculation, uncovering objects that until recently were widely considered to be forever inaccessible to computation. Our computations stem from the "BBP" formula for π, which was discovered in 1997 using a computer program implementing the "PSLQ" integer relation algorithm. This formula has the remarkable property that it permits one to directly calculate binary digits of π, beginning at an arbitrary position d, without needing to calculate any of the first $d-1$ digits. Since 1997, numerous other BBP-type formulas have been discovered for various mathematical constants, including formulas for π^2 (both in binary and ternary bases), and for Catalan's constant.

In this article we describe the computation of base-64 digits of π^2, base-729 digits of π^2, and base-4096 digits of Catalan's constant, in each case beginning at the ten trillionth place, computations that involved a total of approximately 1.549×10^{19} floating-point operations. We also discuss connections between BBP-type formulas and the age-old unsolved questions of whether and why constants such as $\pi, \pi^2, \log 2$ and Catalan's constant have "random" digits.

2 Historical background

Since the dawn of civilization, mathematicians have been intrigued by the digits of π [6], more so than any other mathematical constant. In the third century BCE, Archimedes employed a brilliant scheme of inscribed and circumscribed $3 \cdot 2^n$-gons to compute π to two decimal digit accuracy. However, this and other numerical calculations of antiquity were severely hobbled by their reliance on primitive arithmetic systems.

*Lawrence Berkeley National Laboratory, Berkeley, CA 94720, dhbailey@lbl.gov. Supported in part by the Director, Office of Computational and Technology Research, Division of Mathematical, Information, and Computational Sciences of the U.S. Department of Energy, under contract number DE-AC02-05CH11231.

†Centre for Computer Assisted Research Mathematics and its Applications (CARMA), University of Newcastle, Callaghan, NSW 2308, Australia, jonathan.borwein@newcastle.edu.au.

‡IBM Australia, 601 Pacific Highway, St. Leonards, NSW 2065, Australia, andrew_mattingly@au1.ibm.com.

§IBM Research, 55 Pyrmont Road, Pyrmont, NSW 2009, Australia, glenn_wightwick@au1.ibm.com.

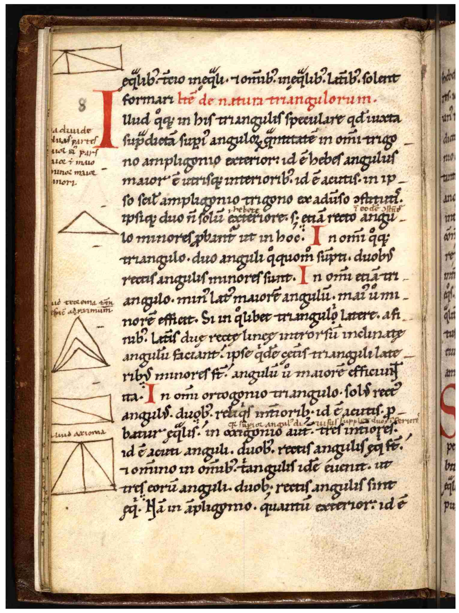

Figure 1: Excerpt from *de Geometria* by Pope Sylvester II (reigned 999-1002 CE)

One of the most significant scientific developments of history was the discovery of full positional decimal arithmetic with zero, by an unknown mathematician or mathematicians in India at least by 500 CE, and probably earlier. Some of the earliest documentation includes the *Araybhatiya*, the writings of the Indian mathematician Araybhata dated to 499 CE, the *Lokavibhaga*, a cosmological work with astronomical observations that permit modern scholars to conclude that it was written on 25 August 458 CE [9], and the Bakhshali manuscript, an ancient mathematical treatise that some scholars believe may be older still, but in any event is no later than the seventh century [7, 8, 2]. The Bakhshali manuscript includes, among other things, the following intriguing algorithm for computing the square root of q, starting with an approximation x_0:

$$\begin{aligned} a_n &= \frac{q - x_n^2}{2x_n} \\ x_{n+1} &= x_n + a_n - \frac{a_n^2}{2\,(x_n + a_n)}. \end{aligned} \tag{1}$$

This scheme is quartically convergent, in that it approximately *quadruples* the number of correct digits with each iteration (although it was never iterated more than once in the examples given in the manuscript) [2].

In the 10th century, Gerbert of Aurillac, who later reigned as Pope Sylvester II, attempted to introduce decimal arithmetic in Europe, but little headway was made until the publication of Fibonacci's *Liber Abaci* in 1202. Several hundred more years would pass before the system finally gained universal, if belated, adoption in the West. The time of Sylvester's reign was a very turbulent one, and he died in 1002, shortly after the death of his protector, Emperor Otto III. It is interesting to speculate how history would have changed had he lived longer. A page from his mathematical treatise *De Geometria* is shown in Figure 1.

Figure 2: The ENIAC in the Smithsonian Museum

2.1 The age of Newton

Armed with decimal arithmetic, and spurred by the newly discovered methods of calculus, mathematicians computed with aplomb. Again, the numerical value of π was a favorite target. Isaac Newton devised an arcsine-like scheme to compute digits of π and recorded 15 digits, although he sheepishly acknowledged, "I am ashamed to tell you to how many figures I carried these computations, having no other business at the time." Newton wrote these words during the plague year 1666, when, ensconced in a country estate, he devised the fundamentals of calculus and the laws of optics, motion and gravitation.

All large computations of π until 1980 relied on variations of Machin's formula:

$$\frac{\pi}{4} \;=\; 4\arctan\left(\frac{1}{5}\right) - \arctan\left(\frac{1}{239}\right). \tag{2}$$

The culmination of these feats was a computation of π using (2) to 527 digits in 1853 by William Shanks, later (erroneously) extended to 607 and then 707 digits. In the preface to the publication of this computation, Shanks wrote that his work "would add little or nothing to his fame as a Mathematician, though it might as a Computer" (until 1950 the word "computer" was used for a person, and the word "calculator" was used for a machine).

One motivation for such computations was to see whether the digits of π repeat, thus disclosing the fact that π is a ratio of two integers. This was settled in 1766, when Lambert proved that π is irrational, thus establishing that the digits of π do not repeat in any number base. In 1882, Lindemann established that π is transcendental, thus establishing that the digits of π^2 or any integer polynomial of π cannot repeat, and also settling once and for all the ancient Greek question of whether the circle could be squared — it cannot, because all numbers that can be formed by finite straightedge-and-compass constructions are necessarily algebraic.

2.2 The computer age

At the dawn of the computer age, John von Neumann suggested computing digits of prominent mathematical constants, including π and e, for statistical analysis. At his instigation,

π was computed to 2037 digits in 1949 on the *Electronic Numerical Integrator And Calculator* (ENIAC) — see Figure 2. In 1965 mathematicians realized that the newly-discovered fast Fourier transform could be used to dramatically accelerate high-precision multiplication, thus facilitating not only large calculations of π and other mathematical constants, but research in computational number theory as well.

In 1976, Eugene Salamin and Richard Brent independently discovered new algorithms for computing the elementary exponential and trigonometric functions (and thus constants such as π and e) much more rapidly than by using classical series expansions. Their schemes, based on elliptic integrals and the Gauss arithmetic-geometric mean iteration, approximately *double* the number of correct digits in the result with each iteration. Armed with such techniques, π was computed to over one million digits in 1973, to over one billion digits in 1989, to over one trillion digits in 2002, and to over five trillion digits at the present time— see Table 1.

Name	Year	Correct Digits
Miyoshi and Kanada	1981	2,000,036
Kanada-Yoshino-Tamura	1982	16,777,206
Gosper	1985	17,526,200
Bailey	Jan. 1986	29,360,111
Kanada and Tamura	Sep. 1986	33,554,414
Kanada and Tamura	Oct. 1986	67,108,839
Kanada et. al	Jan. 1987	134,217,700
Kanada and Tamura	Jan. 1988	201,326,551
Chudnovskys	May 1989	480,000,000
Kanada and Tamura	Jul. 1989	536,870,898
Kanada and Tamura	Nov. 1989	1,073,741,799
Chudnovskys	Aug. 1991	2,260,000,000
Chudnovskys	May 1994	4,044,000,000
Kanada and Takahashi	Oct. 1995	6,442,450,938
Kanada and Takahashi	Jul. 1997	51,539,600,000
Kanada and Takahashi	Sep. 1999	206,158,430,000
Kanada-Ushiro-Kuroda	Dec. 2002	1,241,100,000,000
Takahashi	Jan. 2009	1,649,000,000,000
Takahashi	Apr. 2009	2,576,980,377,524
Bellard	Dec. 2009	2,699,999,990,000
Kondo and Yee	Aug. 2010	5,000,000,000,000

Table 1: Modern computer-era π calculations

Similarly, the constants $e, \phi = \frac{1+\sqrt{5}}{2}, \sqrt{2}, \log 2, \log 10, \zeta(3)$, Catalan's constant $G = \sum_{n=0}^{\infty}(-1)^n/(2n+1)^2$, and Euler's γ constant have now been computed computed to impressive numbers of digits — see Table 2 [10]. For Euler's constant the most efficient method is due to Brent and McMillan (the Nobel physicist) and relies on approximating γ by the ratio of Bessel function values [6], while the log constants are best computed as approximations of elliptic integrals [6].

One of the most intriguing aspects of this historical chronicle is the repeated assurances, often voiced by highly knowledgeable people, that future progress would be limited. As recently h

Constant	Decimal digits	Researcher	Date
$\sqrt{2}$	1,000,000,000,000	S. Kondo	2010
ϕ	1,000,000,000,000	A. Yee	2010
e	500,000,000,000	S. Kondo	2010
$\log 2$	100,000,000,000	S. Kondo	2011
$\log 10$	100,000,000,000	S. Kondo	2011
$\zeta(3)$	100,000,001,000	A. Yee	2011
G	31,026,000,000	A. Yee and R. Chan	2009
γ	29,844,489, 545	A. Yee	2010

Table 2: Computations of other mathematical constants

as 1963, Daniel Shanks, who himself calculated π to over 100,000 digits, told Philip Davis that computing one billion digits would be "forever impossible." Yet this feat was achieved less than 30 years later in 1989 by Yasumasa Kanada of Japan. Also, in 1989, famous British physicist Roger Penrose, in the first edition of his best-selling book *The Emperor's New Mind*, declared that humankind likely will never know if a string of ten consecutive sevens occurs in the decimal expansion of π. This string was found just eight years later, in 1997, also by Kanada, beginning at position 22,869,046,249. After being advised of this fact by one of the present authors, Penrose revised his second edition to specify twenty consecutive sevens.

Along this line, Brouwer and Heyting, exponents of the "intuitionist" school of mathematical logic, proposed, as a premier example of a hypothesis that could never be formally settled, the question of whether and when the string "0123456789" appears in the decimal expansion of π. Kanada found this at the 17,387,594,880-th position after the decimal point. Even astronomer Carl Sagan, whose lead character in his 1985 novel *Contact* (played by Jodi Foster in the movie version) sought confirmation in base-11 digits of π, expressed surprise to learn, shortly after the book's publication, that π had already been computed to many millions of digits.

3 The BBP formula for pi

A 1997 paper [3], [5, Ch. 3] by one of the present authors (Bailey), Peter Borwein and Simon Plouffe presented the following previously unknown formula for π, now known as the "BBP" formula for π:

$$\pi = \sum_{k=0}^{\infty} \frac{1}{16^k} \left(\frac{4}{8k+1} - \frac{2}{8k+4} - \frac{1}{8k+5} - \frac{1}{8k+6} \right). \tag{3}$$

This formula has the remarkable property that it permits one to directly calculate binary or hexadecimal digits of π beginning at an arbitrary starting position, without needing to calculate any of the preceding digits. The resulting simple algorithm requires only minimal memory, does not require multiple-precision arithmetic, and is very well suited to highly parallel computation. The cost of this scheme increases only slightly faster than the index of the starting position.

The proof of this formula is surprisingly elementary. First note that for any $k < 8$,

$$\int_0^{1/\sqrt{2}} \frac{x^{k-1}}{1-x^8}\,\mathrm{d}x \;=\; \int_0^{1/\sqrt{2}} \sum_{i=0}^{\infty} x^{k-1+8i}\,\mathrm{d}x \;=\; \frac{1}{2^{k/2}} \sum_{i=0}^{\infty} \frac{1}{16^i(8i+k)}. \tag{4}$$

Thus one can write

$$\sum_{i=0}^{\infty} \frac{1}{16^i}\left(\frac{4}{8i+1} - \frac{2}{8i+4} - \frac{1}{8i+5} - \frac{1}{8i+6}\right) \;=\; \int_0^{1/\sqrt{2}} \frac{4\sqrt{2} - 8x^3 - 4\sqrt{2}x^4 - 8x^5}{1-x^8}\,\mathrm{d}x, \tag{5}$$

which on substituting $y := \sqrt{2}x$ becomes

$$\int_0^1 \frac{16\,y - 16}{y^4 - 2\,y^3 + 4\,y - 4}\,\mathrm{d}y \;=\; \int_0^1 \frac{4y}{y^2-2}\,\mathrm{d}y - \int_0^1 \frac{4y-8}{y^2-2y+2}\,\mathrm{d}y \;=\; \pi, \tag{6}$$

reflecting a partial fraction decomposition of the integral on the left-hand side. In 1997 neither *Maple* nor *Mathematica* could evaluate (3) symbolically to produce the result π. Today both systems can do this easily.

3.1 Binary digits of log 2

It is worth noting that the BBP formula (3) was not discovered by a conventional analytic derivation. Instead, it was discovered via a computer-based search using the PSLQ *integer relation detection algorithm* (see Section 3.2) of mathematician-sculptor Helaman Ferguson [4], in a process that some have described as an exercise in "reverse mathematical engineering." The motivation for this search was the earlier observation by the authors of [3] that $\log 2$ also has this arbitrary position digit calculating property. This can be seen by analyzing the classical formula

$$\log 2 \;=\; \sum_{k=1}^{\infty} \frac{1}{k2^k}, \tag{7}$$

which has been known at least since the time of Euler, and which is closely related to the functional equation for the dilogarithm.

Let $r \bmod 1$ denote the fractional part of a nonnegative real number r, and let d be a nonnegative integer. Then the binary fraction of $\log 2$ after the "decimal" point has been shifted to the right d places can be written as

$$\begin{aligned}(2^d \log 2) \bmod 1 \;&=\; \left(\sum_{k=1}^{d} \frac{2^{d-k}}{k} \bmod 1 + \sum_{k=d+1}^{\infty} \frac{2^{d-k}}{k} \bmod 1\right) \bmod 1 \\ &=\; \left(\sum_{k=1}^{d} \frac{2^{d-k} \bmod k}{k} \bmod 1 + \sum_{k=d+1}^{\infty} \frac{2^{d-k}}{k} \bmod 1\right) \bmod 1, \end{aligned} \tag{8}$$

where "mod k" has been inserted in the numerator of first term since we are only interested in the fractional part of the result after division.

The operation $2^{d-k} \bmod k$ can be performed very rapidly by means of the *binary algorithm for exponentiation.* This scheme is the simple observation that an exponentiation operation such as 3^{17} can be performed in only five multiplications, instead of 16, by writing it as $3^{17} = ((((3^2)^2)^2)^2) \cdot 3$. Additional savings can be realized by reducing all of the intermediate multiplication results modulo k at each step. This algorithm, together with the division and summation operations indicated in the first term, can be performed in ordinary double-precision floating-point arithmetic, or, for very large calculations by using quad- or oct-precision arithmetic.

Expressing the final fractional value in binary notation yields a string of digits corresponding to the binary digits of $\log 2$ beginning immediately after the first d digits of $\log 2$. Computed results can be easily checked by performing this operation for two slightly different positions, say $d-1$ and d, then checking to see that resulting digit strings properly overlap.

3.2 Hunt for a pi formula

In the wake of finding the above scheme for the binary digits of $\log 2$, the authors of [3] immediately wondered if there was a similar formula for π (none was known at the time). Their approach was to collect a list of mathematical constants (α_i) for which formulas similar in structure to the formula for $\log 2$ were known in the literature, and then to determine, by means of the PSLQ integer relation algorithm, if a nontrivial linear relation exists of the form

$$a_0\pi + a_1\alpha_1 + a_2\alpha_2 + \cdots + a_n\alpha_n = 0, \tag{9}$$

where a_i are integers (because such a relation could then be solved for π to yield the desired formula). After several months of false starts, the following relation was discovered:

$$\pi \;=\; 4 \cdot {}_2F_1\left(\left.\begin{matrix}1, \frac{1}{4}\\ \frac{5}{4}\end{matrix}\right| -\frac{1}{4}\right) + 2\arctan\left(\frac{1}{2}\right) - \log 5, \tag{10}$$

where the first term is a Gauss hypergeometric function evaluation. After writing this formula explicitly in terms of summations, the BBP formula for π was uncovered:

$$\pi = \sum_{k=0}^{\infty} \frac{1}{16^k}\left(\frac{4}{8k+1} - \frac{2}{8k+4} - \frac{1}{8k+5} - \frac{1}{8k+6}\right). \tag{11}$$

One question that immediately arose in the wake of the discovery of the BBP formula for π was whether there are formulas of this type for π in other number bases — in other words, formulas where the 16 in the BBP formula is replaced by some other integer, such as 3 or 10. These computer searches were largely laid to rest in 2004, when one of the present authors (Jonathan Borwein), together with Will Galway and David Borwein showed that there are no degree-1 BBP-type formulas of *Machin-type* for π, except those whose base is a power of two [5, pg. 131–133].

3.3 The BBP formula in action

Variants of the BBP formula have been used in numerous computations of high-index digits of π. In 1998 Colin Percival, then a 17-year-old undergraduate at Simon Fraser University in

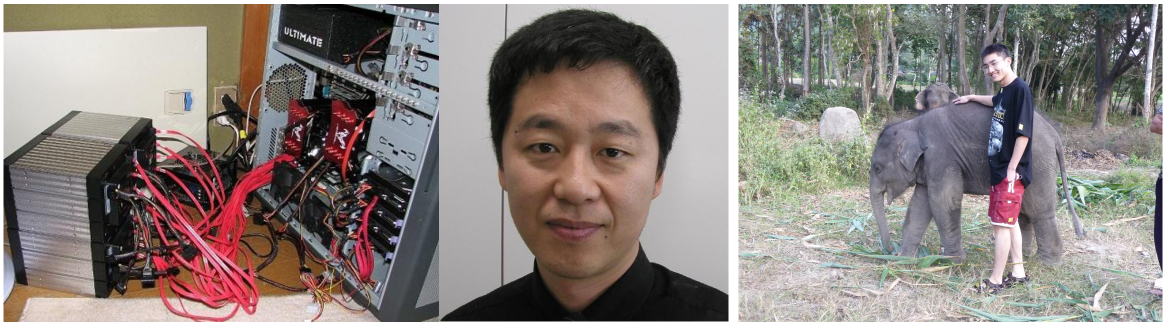

Figure 3: (L) Shigeru Kondo and his π-computer. (R) Alex Yee and his elephant

Canada, computed binary digits beginning at position one quadrillion (10^{15}). At the time, this was one of the largest, if not the largest, distributed computations ever done. More recently, in July 2010, Tsz-Wo Sze of *Yahoo! Cloud Computing*, in roughly 500 CPU-years of computing on *Apache Hadoop* clusters, found that the base-16 digits of π beginning at position 5×10^{14} (corresponding to binary position two quadrillion) are:

`0 E6C1294A ED40403F 56D2D764 026265BC A98511D0 FCFFAA10 F4D28B1B B5392B8.`

The BBP formulas have also been used to confirm other computations of π. For example, in August 2010, Shigeru Kondo (a hardware engineer) and Alexander Yee (an undergraduate software engineer) computed five trillion decimal digits of π on a home-built $18,000 machine. They found that the last 30 digits leading up to position five trillion are

`7497120374 4023826421 9484283852.`

Kondo and Yee (see photos in Figure 3) used the following Chudnovsky-Ramanujan series:

$$\frac{1}{\pi} \quad = \quad 12 \sum_{k=0}^{\infty} \frac{(-1)^k \, (6k)!(13591409 + 545140134k)}{(3k)! \, (k!)^3 \, 640320^{3k+3/2}}, \tag{12}$$

They did not merely evaluate this formula as written, but instead employed a clever quasi-symbolic scheme that mostly avoids the need for full-precision arithmetic.

Kondo and Yee first computed their result in hexadecimal (base-16) digits. Then, in a crucial verification step, they checked hex digits near the end against the same string of digits computed using the BBP formula for π. When this test passed, they converted their entire result to decimal. The entire computation took 90 days, including 64 hours for the BBP confirmation and 8 days for base conversion to decimal. Note that the much lower time for the BBP confirmation, relative to the other two parts, greatly reduced the overall computational cost. A description of their work is available at [11].

4 BBP-type formulas for other constants

In the years since 1997, computer searches using the PSLQ algorithm, as well as conventional analytic investigations, have uncovered BBP-type formulas for numerous other mathematical constants, including $\pi^2, \log^2 2, \pi \log 2,\ \zeta(3), \pi^3, \log^3 2, \pi^2 \log 2, \pi^4, \zeta(5)$ and Catalan's constant. BBP formulas are also known for many arctangents, and for $\log k,\ 2 \le k \le 22$, although none is known for $\log 23$. These formulas and many others, together with references, are given in an online compendium [1].

One particularly intriguing fact is that whereas only binary formulas exist for π, there are both binary and ternary (base-3) formulas for π^2:

$$\pi^2 = \frac{9}{8}\sum_{k=0}^{\infty}\frac{1}{64^k}\left(\frac{16}{(6k+1)^2}-\frac{24}{(6k+2)^2}-\frac{8}{(6k+3)^2}-\frac{6}{(6k+4)^2}+\frac{1}{(6k+5)^2}\right). \tag{13}$$

$$\begin{aligned}\pi^2 = \frac{2}{27}\sum_{k=0}^{\infty}\frac{1}{729^k}&\left(\frac{243}{(12k+1)^2}-\frac{405}{(12k+2)^2}-\frac{81}{(12k+4)^2}-\frac{27}{(12k+5)^2}\right.\\ &\left.-\frac{72}{(12k+6)^2}-\frac{9}{(12k+7)^2}-\frac{9}{(12k+8)^2}-\frac{5}{(12k+10)^2}+\frac{1}{(12k+11)^2}\right).\end{aligned} \tag{14}$$

Formula (13) appeared in [3], while formula (14) is due to Broadhurst. There are known binary BBP formulas for both $\zeta(3)$ and π^3, but no one has found a ternary formula for either.

4.1 Catalan's constant

One other mathematical constant of central interest is Eugéne Charles *Catalan's* (1814-1894) constant

$$G = \sum_{n=0}^{\infty}\frac{(-1)^n}{(2n+1)^2} = 0.91596559417722\ldots, \tag{15}$$

which is arguably the most basic constant whose irrationality and transcendence (though strongly suspected) remain unproven. Note the close connection to this formula for π^2:

$$\frac{\pi^2}{8} = \sum_{n=0}^{\infty}\frac{1}{(2n+1)^2} = 1.2337005501362\ldots. \tag{16}$$

Formulas (15) and (16) can be viewed as the simplest Dirichlet L-series values at 2. Such considerations were behind our decision to focus the computation described in this paper on these two constants.

Catalan's constant has already been the subject of large computations. As mentioned above, in 2009 Alexander Yee and Raymond Chan calculated G to 31.026 billion digits [10]. This computation employed two formulas, including this formula due to Ramanujan:

$$G = \frac{3}{8}\sum_{n=0}^{\infty}\frac{1}{\binom{2n}{n}(2n+1)^2} + \frac{\pi}{8}\log(2+\sqrt{3}), \tag{17}$$

which can be derived from the fact that $G = -T(\pi/4) = -3/2 \cdot T(\pi/12)$, where $T(\theta) := \int_0^\theta \log\tan\sigma\,\mathrm{d}\sigma$.

The BBP compendium lists two BBP-type formulas for G. The first was discovered numerically by Bailey, but both it and the second formula were subsequently proven by Kunle Adegoke, based in part on some results of Broadhurst.

For the present study, we sought a formula for G with as few terms as possible, because the run time for computing with a BBP-type formula increases roughly linearly with the number of nonzero coefficients. The two formulas in the compendium have 22 and 18 nonzero coefficients, respectively. So we explored, by means of a computation involving the PSLQ algorithm, the linear space of formulas for G spanned by these two sets of coefficients, together with two known "zero relations" (BBP-type formulas whose sum is zero). These analyses and computations led to the following formula, which has only 16 nonzero coefficients, and which we believe to be the most economical BBP-type formula for computing Catalan's constant:

$$\begin{aligned} G &= \frac{1}{4096}\sum_{k=0}^{\infty}\frac{1}{4096^k}\left(\frac{36864}{(24k+2)^2}-\frac{30720}{(24k+3)^2}-\frac{30720}{(24k+4)^2}-\frac{6144}{(24k+6)^2}-\frac{1536}{(24k+7)^2}\right. \\ &\quad +\frac{2304}{(24k+9)^2}+\frac{2304}{(24k+10)^2}+\frac{768}{(24k+14)^2}+\frac{480}{(24k+15)^2}+\frac{384}{(24k+11)^2}+\frac{1536}{(24k+12)^2} \\ &\quad \left.+\frac{24}{(24k+19)^2}-\frac{120}{(24k+20)^2}-\frac{36}{(24k+21)^2}+\frac{48}{(24k+22)^2}-\frac{6}{(24k+23)^2}\right). \end{aligned} \tag{18}$$

5 BBP formulas and normality

One prime motivation in computing and analyzing digits of π and other well-known mathematical constants through the ages is to explore the age-old question of whether and why these digits appear "random." Numerous computer-based statistical checks of the digits of π — unlike those of e — so far have failed to disclose any deviation from reasonable statistical norms. See, for instance, Table 3, which presents the counts of individual hexadecimal digits among the first trillion hex digits, as obtained by Yasumasa Kanada.

Given some positive integer b, a real number α is said to be b-normal if every m-long string of base-b digits appears in the base-b expansion of α with precisely the expected limiting frequency $1/b^m$. It follows from basic probability theory that almost all real numbers are b-normal for any specific base b and even for all bases simultaneously. But proving normality for specific constants of interest in mathematics has proven remarkably difficult.

Interest in BBP-type formulas was heightened by the 2001 observation, by one of the present authors (Bailey) and Richard Crandall, that the normality of BBP-type constants such as $\pi, \pi^2, \log 2$ and G can be reduced to a certain hypothesis regarding the behavior of a class of chaotic iterations [5, pg. 141–173]. No proof is known for this genesis hypothesis, but even specific instances of this result would be quite interesting. For example, if it could be established that the iteration given by $w_0 = 0$, and

$$w_n = \left(2w_{n-1} + \frac{1}{n}\right) \bmod 1 \tag{19}$$

is equidistributed in $[0, 1)$ (i.e., is a "good" pseudorandom number generator), then, according to the Bailey-Crandall result, it would follow that $\log 2$ is 2-normal. In a similar vein, if it could be established that the iteration given by $x_0 = 0$ and

$$x_n = \left(16x_{n-1} + \frac{120n^2 - 89n + 16}{512n^4 - 1024n^3 + 712n^2 - 206n + 21}\right) \bmod 1 \tag{20}$$

is equidistributed in $[0, 1)$, then it would follow that π is 2-normal.

Hex Digit	Occurrences
0	62499881108
1	62500212206
2	62499924780
3	62500188844
4	62499807368
5	62500007205
6	62499925426
7	62499878794
8	62500216752
9	62500120671
A	62500266095
B	62499955595
C	62500188610
D	62499613666
E	62499875079
F	62499937801
Total	1000000000000

Table 3: Digit counts in the first trillion hexadecimal (base-16) digits of π. Note that deviations from the average value 62,500,000,000 occur only after the first six digits, as expected.

Giving further hope to these studies is the recent extension of these methods to a rigorous proof of normality for an uncountably infinite class of real numbers. Given a real number r in $[0,1)$, let r_k denote the k-th binary digit of r. Then the real number

$$\alpha_{2,3}(r) \quad = \quad \sum_{k=0}^{\infty} \frac{1}{3^k 2^{3^k + r_k}} \tag{21}$$

is 2-normal. For example, the constant $\alpha_{2,3}(0) = \sum_{k \geq 0} 1/(3^k 2^{3^k}) = 0.541883680831502985\ldots$ is provably 2-normal. A similar result applies if 2 and 3 in this formula are replaced by any pair of co-prime integers (b,c) greater than one [5, pg. 141–173].

5.1 A curious hexadecimal conjecture

It is tantalizing that if, using (20), one calculates the hexadecimal digit sequence

$$y_n = \lfloor 16 x_n \rfloor \tag{22}$$

(where $\lfloor \cdot \rfloor$ denotes greatest integer), then the sequence (y_n) appears to perfectly (not just approximately) produce the hexadecimal expansion of π. In explicit computations, we checked that the first 10,000,000 hexadecimal digits generated by this sequence are *identical* with the first 10,000,000 hexadecimal digits of $\pi - 3$. This is a fairly difficult computation, as it requires roughly n^2 bit-operations, and is not easily performed on a parallel computer system. In our implementation, computing $2,000,000$ hex digits with (22), using *Maple*, required

17.3 hours on a laptop. Computing 4,100,000 using *Mathematica*, with a more refined implementation, required 46.5 hours. The full confirmation, using a C++ program, took 433,192 seconds (120.3 hours) on a IBM Power 780 (model: 9179-MHB, clock speed: 3.864 GHz). All these outputs were confirmed against stored hex digits of π in the software section of `http://www.experimentalmath.info`.

Conjecture 1 *The sequence $\lfloor 16x_n \rfloor$, where (x_n) is the sequence of iterates defined in equation (20), generates precisely the hexadecimal expansion of $\pi - 3$.*

We can learn more. Let $||x-y|| = \min(|x-y|, |1-(x-y)|)$ denote the "wrapped" distance between reals x and y in $[0,1)$. The base-16 expansion of π, which we denote π_n, satisfies

$$||\pi_n - x_n|| \le \sum_{k=n+1}^{\infty} \frac{120k^2 - 89k + 16}{16^{k-n}(512k^4 - 1024k^3 + 712k^2 - 206k + 21)} \approx \frac{1}{64(n+1)^2}, \tag{23}$$

so that, upon summing from some N to infinity, we obtain the finite value

$$\sum_{n=N}^{\infty} ||\pi_n - x_n|| \ \le \ \frac{1}{64(N+1)}. \tag{24}$$

Heuristically, let us assume that the π_n are independent, uniformly distributed random variables in $(0,1)$, and let $\delta_n = ||\alpha_n - x_n||$. Note that an error (i.e., an instance where x_n lies in a different subinterval of the unit interval than π_n, so that the corresponding hex digits don't match) can only occur when π_n is within δ_n of one of the points $(0, 1/16, 2/16, \cdots, 15/16)$. Since $x_n < \pi_n$ for all n (where $<$ is interpreted in the wrapped sense when x_n is slightly less than one), this event has probability $16\delta_n$. Then the fact that the sum (24) has a finite value implies, by the first *Borel-Cantelli* lemma, that there can only be finitely many errors. Further, the small value of the sum (24), even when $N = 1$, suggests that it is unlikely that any errors will be observed. If we set $N = 10,000,001$ in (24), since we know there are no errors in the first 10,000,000 elements, then we obtain an upper bound of 1.563×10^{-9} which suggests it is truly unlikely that errors will ever occur.

A similar correspondence can be seen between iterates of (19) and the binary digits of $\log 2$. In particular, let $z_n = \lfloor 2w_n \rfloor$, where w_n is given in (19). Then since the sum of the error terms for $\log 2$, corresponding to (24), is infinite, it follows by the second Borel-Cantelli lemma that discrepancies between (z_n) and the binary digits of $\log 2$ can be expected to appear indefinitely, but with decreasing frequency. Indeed, in computations that we have done, we have found that the sequence (z_n) disagrees with 10 of the first 20 binary digits of $\log 2$, but in only one position over the range 5000 to 8000.

6 Computing digits of π^2 and Catalan's constant

In illustration of this theory, we now present the results of computations of high-index binary digits of π^2, ternary digits of π^2, and binary digits of Catalan's constant, based on formulas (13), (14) and (18), respectively. These calculations were performed on a 4-rack *BlueGene/P* system at IBM's Benchmarking Centre in Rochester, Minnesota, USA. This is a shared facility,

Figure 4: Andrew Mattingly, Blue Gene/P, and Glenn Wightwick

so calculations were conducted over a several month period, where, at any given time, none, some or all of the system was available. It was programmed remotely from Australia, which permitted the system to be used off-hours. Sometimes it helps to be in a different time zone!

1. *Base-64 digits of π^2 beginning at position 10 trillion.* The first run, which produced base-64 digits starting from position $10^{12} - 1$, required an average of 253,529 seconds per thread, and was subdivided into seven partitions of 2048 threads each, so the total cost was $7 \cdot 2048 \cdot 253529 = 3.6 \times 10^9$ CPU-seconds. Each rack of the IBM system features 4096 cores, so the total cost is 10.3 "rack-days." The second run, which produced base-64 digits starting from position 10^{12}, completed in nearly the same run time (within a few minutes). The two resulting base-8 digit strings are

   ```
   75|6011450530323647572450000574326275453036305241635063|573227604
     |6011450530323647572450000574326275453036305241635063|220210566
   ```

   (each pair of base-8 digits corresponds to a base-64 digit). Here the digits in agreement are delimited by |. Note that 53 consecutive base-8 digits (or, equivalently, 159 consecutive binary digits) are in perfect agreement.

2. *Base-729 digits of π^2 beginning at position 10 trillion.* In this case, the two runs each required an average of 795,773 seconds per thread, similarly subdivided as above, so that the total cost was 6.5×10^9 CPU-seconds, or 18.4 "rack-days." The two resulting base-9 digit strings are

   ```
   001|12264485064548583177111135210162856048323453468|10565567635862
      |12264485064548583177111135210162856048323453468|04744867134524
   ```

   (each triplet of base-9 digits corresponds to one base-729 digit). Note here that 47 consecutive base-9 digits (94 consecutive base-3 digits) are in perfect agreement.

3. *Base-4096 digits of Catalan's constant beginning at position 10 trillion.* These two runs each required 707,857 seconds per thread, but in this case were subdivided into eight partitions of 2048 threads each, so that the total cost was 1.2×10^{10} CPU-seconds, or 32.8 "rack-days." The two resulting base-8 digit strings are

CONSTANT	n'	d	#ITERS ($\times 10^{15}$)	TIME/ITER (microsec)	TIME (yr)	WITH VERIFY	TOTAL (yr)	O'HEAD (%)	FLOPS ($\times 10^{18}$)
π^2 base-2^6	5	10^{13}	2.16	1.424	97.43	194.87	230.35	18.2	2.58
π^2 base-3^6	9	10^{13}	3.89	1.424	175.38	350.76	413.16	17.8	4.65
G base-4^6	16	10^{13}	6.91	1.424	311.79	623.58	735.02	17.9	8.26

Table 4: The scale of our computations. We estimate 4.5 quad-double operations per iteration and that each costs 266 single-precision operations. The total cost in single-precision operations is given in the last column. This total includes overhead which is largely due to a rounding operation that we implemented using bit-masking.

Digit	0	1	2	3	4	5	6	7
base-2 (141)	0.454	0.546	-	-	-	-	-	-
base-4 (70)	0.171	0.329	0.229	0.271	-	-	-	-
base-8 (47)	0.085	0.128	0.213	0.128	0.064	0.128	0.043	0.213

Table 5: Base-4096 digits of G beginning at position 10 trillion: digit proportions

```
0176|34705053774777051122613371620125257327217324522|6000177545727
    |34705053774777051122613371620125257327217324522|5703510516602
```

(each quadruplet of base-8 digits corresponds to one base-4096 digit). Note that 47 consecutive base-8 digits (141 consecutive binary digits) are in perfect agreement.

These long strings of consecutively agreeing digits, beginning with the target digit, provide a compelling level of statistical confidence in the results. In the first case, for instance, note that the probability that 32 pairs of randomly chosen base-8 digits are in perfect agreement is roughly 1.2×10^{-29}. Even if one discards, say, the final six base-8 digits as a 1-in-262,144 statistical safeguard against numerical round-off error, one would still have 24 consecutive base-8 digits in perfect agreement, with a corresponding probability of 2.1×10^{-22}. Now strictly speaking, one cannot define a valid probability measure on digits of π^2, but nonetheless, from a practical point of view, such analysis provides a very high level of statistical confidence that the results have been correctly computed.

For this reason, computations of π and the like are a favorite tool for the integrity testing for computer system hardware and software. If either run of a paired computation of π succumbs to even a single fault in the course of the computation, then typically the final results will disagree almost completely. For example, in 1986, a similar pair of computations of π disclosed some subtle but substantial hardware errors in an early model of the Cray-2 supercomputer. Indeed, the calculations we have done arguably constitute the most strenuous integrity test ever performed on the BlueGene/P system. Table 4 gives some sense of the scale of the three record computations, which used more than 135 "rack-days," 1378 serial CPU-years and more than 1.549×10^{19} floating point operations. This is comparable to the cost of the most sophisticated animated movies as of the present time (2011).

For the sake of completeness, in Table 5 we also record the one, two and three-bit frequency counts from our Catalan computation.

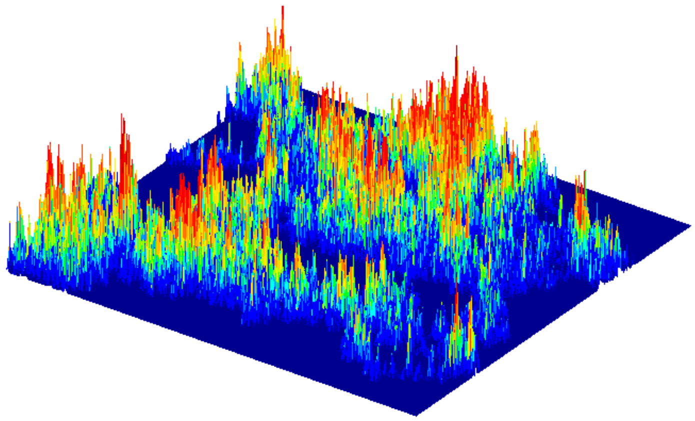

Figure 5: A random walk on a million digits of Catalan's constant

7 Future directions

It is ironic that in an age when even pillars such as Fermat's Last Theorem and the Poincairé conjecture have succumbed to the brilliance of modern mathematics, that one of the most elementary mathematical hypotheses, namely whether (and why) the digits of π or other constants, such as $\log 2, \pi^2$ or G (see Figure 5), are "random," remains unanswered. In particular, proving that π (or $\log 2, \pi^2$ or G) is b-normal in some integer base b remains frustratingly elusive. Even much weaker results, for instance the simple assertion that a one appears in the binary expansion of π (or $\log 2, \pi^2$ or G) with limiting frequency 1/2 (which assertion has been amply affirmed in numerous computations over the years), remain unproven and largely inaccessible at the present time.

Almost as much ignorance extends to simple algebraic irrationals such as $\sqrt{2}$. In this case it is now known that the number of ones in the first n binary digits of $\sqrt{2}$ must be at least of the order of $\sqrt{n}$, with similar results for other algebraic irrationals [5, pg. 141–173]. But this is a very weak result, given that this limiting ratio is almost certainly 1/2, not only for $\sqrt{2}$ but more generally for all algebraic irrationals.

Nor can we prove much about continued fractions for various constants, except for a few well-known results for special cases such as quadratic irrationals, ratios of Bessel functions, and certain expressions involving exponential functions.

For these reasons, there is continuing interest in the theory of BBP-type constants, since, as mentioned, there is an intriguing connection between BBP-type formulas and certain chaotic iterations that are akin to pseudorandom number generators. If these connections can be strengthened, then perhaps normality proofs could be obtained for a wide range of polylogarithmic constants, possibly including $\pi, \log 2, \pi^2$ and G.

As settings change, so do questions. Until the question of efficient single-digit extraction was asked, our ignorance about such issues was not exposed. The case of the exponential series

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} \tag{25}$$

is illustrative. For present purposes, the convergence rate in (25) is too good.

Conjecture 2 *There is no BBP formula for e. Moreover, there is no way to extract individual digits of e significantly more rapidly than by computing the first n digits.*

The same could be conjectured about other numbers including Euler's constant $\gamma = 0.57721566490153\ldots$. In short, until vastly stronger mathematical results are obtained in this area, there will doubtless be continuing interest in computing digits of these constants. In the present vacuum, that is perhaps all that we can do.

References

[1] David H. Bailey, "BBP-Type Formulas," manuscript, 2011, available at `http://www.davidhbailey.com/dhbpapers/bbp-formulas.pdf`.

[2] David H. Bailey and Jonathan M. Borwein, "Ancient Indian square roots: An exercise in forensic paleo-mathematics," *American Mathematical Monthly*, vol. 119, no. 8 (Oct 2012), pg. 646–657.

[3] David H. Bailey, Peter B. Borwein and Simon Plouffe, "On The Rapid Computation of Various Polylogarithmic Constants," *Mathematics of Computation*, vol. 66, no. 218, 1997, pg. 903–913.

[4] David H. Bailey and David J. Broadhurst, "Parallel Integer Relation Detection: Techniques and Applications," *Mathematics of Computation*, vol. 70, no. 236 (Oct 2000), pg. 1719–1736.

[5] J. M. Borwein and D. H. Bailey, *Mathematics by Experiment: Plausible Reasoning in the 21st Century* A.K. Peters Ltd, 2004. Expanded Second Edition, 2008.

[6] L. Berggren, J. M. Borwein and P. B. Borwein, *Pi: a Source Book,* Springer-Verlag, (1997), (2000), (2004). Fourth edition in preparation, 2011.

[7] Bibhutibhusan Datta, "The Bakhshali Mathematics," *Bulletin of the Calcutta Mathematical Society*, vol. 21, no. 1 (1929), pg. 1–60.

[8] Rudolf Hoernle, *On the Bakhshali Manuscript*, Alfred Holder, Vienna, 1887.

[9] Georges Ifrah, *The Universal History of Numbers: From Prehistory to the Invention of the Computer*, translated by David Vellos, E. F. Harding, Sophie Wood and Ian Monk, John Wiley and Sons, New York, 2000.

[10] Alexander Yee, "Large Computations," 7 Mar 2011, available at `http://www.numberworld.org/nagisa_runs/computations.html`.

[11] Alexander Yee and Shigeru Kondo, "5 Trillion Digits of Pi — New World Record," 7 Mar 2011, available at `http://www.numberworld.org/misc_runs/pi-5t/details.html`.

Acknowledgements Thanks are due to many colleagues, but most explicitly to Prof Mary-Anne Williams of University Technology Sydney who conceived the idea of a π-related computation to conclude in conjunction with a public lecture at UTS on 3.14.2011 (see `http://datasearch2.uts.edu.au/feit/news-events/event-detail.cfm?ItemId=25541`.) We also wish to thank Matthew Tam who constructed the database version of [1].