CHAMPing at the Bit: Short Square-Root Collisions for Two-Generator Cayley Hashes

Freeman Slaughter

University of South Florida, Tampa FL, USA
fslaughter@usf.edu

Abstract. For $A, B \in \mathrm{GL}_2(\mathbb{F}_p)$, a Cayley hash $H_M(w)$ maps a bit string $w \in \{0,1\}^*$ to the corresponding product of A and B. We construct collisions by finding two palindromes whose matrix products commute. A symmetric form shared by the two generators places the hash of every palindrome in a three-dimensional subspace containing the identity. Two such hashes commute when their traceless parts are proportional, which defines a class map with $p+2$ values, so the pigeonhole principle gives collisions of $4\lceil \log_2(p+3) \rceil$ bits. Under a uniformity heuristic, a dictionary search finds such a pair after about $\sqrt{\pi p/2}$ class evaluations. Adding one bit to each half-message gives an expected bound of $O(\sqrt{p})$ evaluations, with no assumption on how the classes are distributed. When the common form is singular, the generators are simultaneously triangular and the same argument applies to ordinary words.

For CHAMP-512 and CHAMP-1024, the heuristic search finds collisions of 512 and 1024 bits after about 2^{64} and 2^{128} class evaluations, and the one-bit-longer, distribution-independent variant finds collisions of 516 and 1028 bits at the same leading cost. This matches the asymptotic cost of an earlier commuting-product search for CHAMP, whose messages have $O(\log^2(p))$ bits, and improves on the bound of about p given by an earlier palindrome restriction. For every pair over an odd prime field, this also shortens the collisions of the general algorithm of Mullan and Tsaban from $O(\log^2(p)/\log(\log(p)))$ to $4\lceil \log_2(p+3) \rceil$ bits. Since the argument applies to every pair A, B, no pair of generators gives a 2×2 Cayley hash over $\mathbb{F}_p$ with collision resistance greater than $O(\sqrt{p})$. We give an explicit proof-of-concept collision for the CHAMP matrices at a 36-bit prime, and prove that every sufficiently short message can be recovered from its digest for the main CHAMP pair. To close, we demonstrate that since reversal and symbol exchange act as sign changes on quaternion coordinates, the alternative CHAMP pair possesses a much greater degree of symmetry, but do not yet see how to utilize this to find collisions below the square-root bound.

This research was partially supported by NSF Grant DMS-2342254.

1 Introduction

A Cayley hash function fixes two elements A, B of a group or semigroup and maps a bit string to the product obtained by replacing each 0 by A and each 1 by B. The idea goes back to Zémor [19] and to Tillich and Zémor [13,14]. Its main attraction is the homomorphic property $H_M(XY) = H_M(X)H_M(Y)$, which makes hashing incremental and parallelizable. Several concrete instantiations of (A, B) have been broken through algebraic structure: integer lifting [13], polynomial methods in characteristic two [9,5], and quaternion methods for LPS graphs [15].

CHAMP [4] takes $A, B \in \mathrm{M}_2(\mathbb{Z})$ of determinant 2, reduced modulo the safe prime $p_{128} = 2^{128} - 15449$ or $p_{256} = 2^{256} - 36113$. Entrywise inversion followed by serialization produces a 512-bit or 1024-bit digest. For messages of a fixed length n, the underlying products have determinant 2^n and occupy at most $p(p^2 - 1)$ matrices. The designers argue that the determinant rules out collisions between messages whose lengths differ by less than $(p-1)/2$, that the negative eigenvalues of the "average matrix" $\frac{1}{2}(A+B)$ defeat the length attack of [2], and that the best known collision attack is a birthday search, whose cost of about $p^{3/2}$ provides the estimates 2^{192} for CHAMP-512 and 2^{384} for CHAMP-1024 [4,18].

CHAMP was submitted to the Next-Generation Commercial Cryptographic Algorithms (NGCC) program of China's Institute of Commercial Cryptography Standards, whose first-round candidates were released in September 2026, and have since been analyzed publicly [18,12]. The program asks for collision security of 256 and 512 bits at the two output sizes [12, finding hash-04-1].

We show that *every* pair (A, B) of invertible 2×2 matrices admits a collision attack of cost $O(\sqrt{p})$ with messages of $O(\log(p))$ bits. The attack reduces collision search to matching projective classes of the hashes of palindromes.

Theorem 1. *Let p be an odd prime and $A, B \in \mathrm{GL}_2(\mathbb{F}_p)$. Suppose an invertible symmetric matrix $S \in \mathrm{M}_2(\mathbb{F}_p)$ satisfies $A^T S = SA$ and $B^T S = SB$, and set $m = \lceil \log_2(p+3) \rceil$. There are distinct $u_1, u_2 \in \{0,1\}^m$ such that, for $P_i = u_i \operatorname{rev}(u_i)$,*

$$P_1P_2 \neq P_2P_1, \quad H_M(P_1P_2) = H_M(P_2P_1).$$

If the classes of distinct sampled half-messages behave as independent uniform values from a set of size $p+2$, a dictionary search finds a pair after an expected $\sqrt{\pi(p+2)/2}\,(1 + o(1))$ evaluations.

A nonzero symmetric form always exists, but it need not be invertible. Section 3 showcases an example and argues that simultaneous triangularization addresses this case. Both CHAMP pairs have invertible common forms at the proposed primes.

Theorem 2. *Both CHAMP pairs admit collisions of 512 bits at p_{128} and 1024 bits at p_{256}. Under the heuristic in Theorem 1, the expected dictionary-search costs are about $1.25 \cdot 2^{64}$ and $1.25 \cdot 2^{128}$ evaluations, respectively. Without that heuristic, a randomized dictionary search finds collisions of 516 bits and 1028 bits in expected $O(\sqrt{p})$ class evaluations, using $O(\sqrt{p})$ expected storage.*

The unconditional bound comes from sampling distinct half-messages from a domain at least twice the size of the class set, which results in a valid birthday estimate, however unevenly the classes may be distributed. Each sampled word costs $O(\log(p))$ field operations, so the expected total is $O(\sqrt{p}\log(p))$. A Pollard-rho implementation involving the original message lengths requires its own random-function heuristic, as explained in Section 4.2.

Theorem 3. *Let $n_0 = \lfloor(\log_2(p) - 1)/3\rfloor$. For the main CHAMP pair, a digest promised to have been generated from a message X of length at most n_0 determines X uniquely among such messages. After decoding the digest by four modular inversions, recovery takes $O(|X| + 1)$ arithmetic operations on integers of $O(\log(p))$ bits.*

The bounds are $n_0 = 42$ for CHAMP-512 and $n_0 = 84$ for CHAMP-1024. The proof needs $2^{3n_0+1} \leq p$, and $p_{256} < 2^{256}$ rules out $n_0 = 85$. Theorem 3 is only concerned with short messages, so does not correspond to a preimage attack on arbitrary messages. However, it additionally shows that the size of A^{-1} and B^{-1} modulo p, upon which the designers' defense against length attacks rests, is not the relevant parameter, since the decoding takes place over the integers.

Relation to previous work. Mullan and Tsaban [7] give heuristic square-root collision algorithms for general generators in $\mathrm{SL}_2(\mathbb{F}_q)$, with length $O(\log(q))$ in even characteristic and $O(\log^2(q)/\log\log(q))$ in the general case. For the CHAMP primes, 2 is a square, so scaling both generators by a square root of 2 moves them into $\mathrm{SL}_2(\mathbb{F}_p)$. Equal-length collisions transfer back unchanged. For a general normalized collision of unequal lengths, this transfer is not automatic, because the two scalar factors differ. The commuting-word construction in [7, Section 4.2] avoids that issue by comparing uv with vu. The square-root exponent is therefore not new in general. Palindromes were also used against the characteristic-two Tillich–Zémor function [5], but through lifting to $\mathbb{F}_2[x]$ rather than through commutation.

For CHAMP itself, the NGCC public assessment [12] contains two closely related findings. Finding hash-04-1 observes that $H_M(M)Q$ is symmetric for every palindrome M, where $Q = \left(\begin{smallmatrix} 9 & 14 \\ 14 & 22 \end{smallmatrix}\right)$ is the adjugate of the form S of Section 3. This is Proposition 1 for the main pair, and a birthday search over the resulting set of about p^2 matrices gives same-length collisions after about p palindromes. Finding hash-04-2 records a commuting-product search built on Idrassi's subgroup search [6]: one birthday stage produces words with upper triangular products, and a second matches the code $(a - d)/b$ to find two that commute. Its projected cost has leading factor 2^{64} for CHAMP-512, with messages of $O(\log^2(p))$ bits, under mixing and word-length assumptions. Lifting of short inputs was also reported in [11].

Reversal identities, palindromes, and commuting products have all been utilized before. Mullan and Tsaban prove a reversal identity for generators with $\det(A_0 - A_1) = 0$ and build palindromic collisions after a first birthday stage into the triangular subgroup [7, Lemma 3.4, Corollary 3.6], and their test for

commuting triangular matrices [7, Proposition 4.2] is the code $(a-d)/b$ used in finding hash-04-2. They also show that palindromic collisions of the form $0v0$, $1v1$ used in [5] cannot occur for odd q [7, Appendix B]; the collisions here have a different form P_1P_2, P_2P_1. Our observation is that every pair of generators has a reversal form (Lemma 2), and that the hashes of palindromes already lie in a space where commutation is a single projective condition. One birthday search therefore suffices for arbitrary generators. This shortens the collisions of the general algorithm of [7] for odd prime p from $O(\log^2(p)/\log\log(p))$ bits to $4\lceil\log_2(p+3)\rceil$ bits, proves that collisions of at most this length exist for every pair of invertible generators, and gives the distribution-independent bound of Proposition 3.

Wang [17] has published explicit CHAMP-512 collisions at the full parameters, confirmed in [12, finding hash-04-6]. The shorter pair consists of 334400-bit messages whose matrix products modulo p_{128} agree – which can be checked directly – while the reported search effort was small: a bridge search over 30325 candidates and a relation search of about six seconds. To the best of our knowledge, no explicit CHAMP-1024 collision has been reported. Our 2^{64} figure therefore concerns short collisions only; with long messages allowed, CHAMP-512 is already broken in practice. For CHAMP-1024, we know of no collision attack below 2^{128}.

Organization. Section 2 defines CHAMP and fixes notation. Sections 3 and 4 develop the reversal form and collision construction. Section 5 gives an explicit collision at a small prime. Section 6 proves short-message recovery, and Section 7 discusses the resulting security bounds. The additional symmetry of the alternative pair and the scalar-order calculations are in Appendices A and B.

2 Preliminaries

Throughout this manuscript, let p be an odd prime and define words to be finite strings over $\{0,1\}$. For a word written as $w = w_1 \cdots w_n$, we write $|w| = n$ for its length, $\mathrm{rev}(w) = w_n \cdots w_1$ for its reversal, and $\tilde{w}$ for the word obtained by swapping the role of the symbols 0 and 1. Given $A, B \in \mathrm{GL}_2(\mathbb{F}_p)$, we set $g_0 = A$, $g_1 = B$, and $H_M(w) = g_{w_1} \cdots g_{w_n}$, where we define the hash of the empty word to be I. A *collision* is a pair of distinct words which hash to the same value of H_M. CHAMP applies an injective map (entrywise inversion, with $0 \mapsto 0$) to $H_M(w)$, so collisions of H_M and of the CHAMP digest are equivalent. The Julia code in [4] writes each inverted entry big-endian, whereas the CHAMP-512 digests published with [17], which were checked against the reference C code, store each entry little-endian. Collisions ultimately do not depend on this choice.

For $X \in \mathrm{M}_2(\mathbb{F}_p)$, write $\mathrm{adj}(X)$ for its classical adjugate, so $X\,\mathrm{adj}(X) = \det(X)I$. With the twist $J = \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$, we have $\mathrm{adj}(X) = J^{-1}X^TJ$. As p is odd, each matrix has a unique decomposition

$$X = \frac{1}{2}\,\mathrm{tr}(X)I + X_0, \quad \mathrm{tr}(X_0) = 0.$$

Cayley–Hamilton gives $X_0^2 = -\det(X_0)I$.

The CHAMP matrices are

$$A = \begin{pmatrix} -2 & 1 \\ 4 & -3 \end{pmatrix}, \quad B = \begin{pmatrix} -5 & 2 \\ -6 & 2 \end{pmatrix}, \quad B' = \begin{pmatrix} 6 & 4 \\ -2 & -1 \end{pmatrix},$$

where (A, B) is the pair used in the CHAMP implementation and (A, B') is the alternative pair. All three have determinant 2. The proposed primes are $p_{128} = 2^{128} - 15449$ and $p_{256} = 2^{256} - 36113$; both are safe primes, $p = 2q + 1$ with q prime, and both satisfy $p \equiv 7 \pmod 8$.

We call a nonempty word w a *scalar word* if $H_M(w) = \lambda I$ for some $\lambda \in \mathbb{F}_p \setminus \{0\}$.

Lemma 1. *Let $p = 2q+1$ be a safe prime with $p \equiv 7 \pmod 8$, and let $\det(A) = \det(B) = 2$. If $H_M(u) = H_M(v)$, then q divides $|u|-|v|$. In particular, $H_M(w) = I$ implies $q \mid |w|$. The determinant alone imposes no restriction on the length of a scalar word when its scalar value is unspecified.*

Proof. The element 2 is a quadratic residue modulo p, so its multiplicative order divides q. Since $2 \neq 1$, this order must be q. Taking determinants proves the first two statements. For the last, the equation $\lambda^2 = 2^{|w|}$ has a solution for every length, since 2 is a square.

The last part of Lemma 1 is important because a scalar word commutes with every word, and therefore produces collisions (Lemma 7). The determinant excludes collisions with $0 < \big||u| - |v|\big| < q$, but does not apply to words of equal length. Scalar words and scalar orders are discussed in more detail in Appendix B.

3 A Common Reversal Form

Lemma 2. *For any $A, B \in \mathrm{M}_2(\mathbb{F})$ over a field $\mathbb{F}$, there exists a nonzero symmetric matrix S such that SA and SB are symmetric.*

Proof. The symmetric 2×2 matrices form a space of dimension 3. For fixed C, the map $S \mapsto (SC)_{12} - (SC)_{21}$ is linear, and SC is symmetric exactly when it vanishes. Two linear conditions on a three-dimensional space leave a solution space of dimension at least 1.

When S is invertible, define $\rho_S(X) = S^{-1}X^TS$. It reverses products and, by symmetry of S, is an involution:

$$\rho_S(XY) = \rho_S(Y)\rho_S(X), \quad \rho_S(\rho_S(X)) = S^{-1}S^TXS^{-T}S = X.$$

Proposition 1. *Let S be invertible and symmetric with SA and SB symmetric. Then for every word w,*

$$H_M(\mathrm{rev}(w)) = S^{-1}\, H_M(w)^T\, S.$$

Proof. If SC is symmetric then $C^T S = C^T S^T = (SC)^T = SC$, so $\rho_S(C) = C$. Hence ρ_S fixes A and B. Since ρ_S reverses products, $\rho_S(g_{w_1} \cdots g_{w_n}) = g_{w_n} \cdots g_{w_1} = H_M(\mathrm{rev}(w))$.

For the two CHAMP pairs, one may take

$$S = \begin{pmatrix} 22 & -14 \\ -14 & 9 \end{pmatrix} \text{ for } (A, B), \quad \widehat{S} = \begin{pmatrix} 10 & 6 \\ 6 & 1 \end{pmatrix} \text{ for } (A, B'),$$

where $\det(S) = 2$ and $\det(\widehat{S}) = -26$. Direct computation then gives

$$SA = \left(\begin{smallmatrix} -100 & 64 \\ 64 & -41 \end{smallmatrix}\right), \quad SB = \left(\begin{smallmatrix} -26 & 16 \\ 16 & -10 \end{smallmatrix}\right), \quad \widehat{S}A = \left(\begin{smallmatrix} 4 & -8 \\ -8 & 3 \end{smallmatrix}\right), \quad \widehat{S}B' = \left(\begin{smallmatrix} 48 & 34 \\ 34 & 23 \end{smallmatrix}\right).$$

Both forms are invertible modulo every odd prime besides 13, so Proposition 1 holds for both CHAMP primes. For the main pair, Proposition 1 appears in [12, finding hash-04-1] in the equivalent form that $H_M(M)\,\mathrm{adj}(S)$ is symmetric for every palindrome M.

Remark 1. Lemma 2 does not necessarily give an *invertible* S. For example, take

$$A_0 = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}, \quad B_0 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

Over any field, the conditions on $S = \left(\begin{smallmatrix} s_{11} & s_{12} \\ s_{12} & s_{22} \end{smallmatrix}\right)$ are equivalent to $2s_{12} = s_{12}$ and $s_{11} + s_{12} = s_{12}$, thus $s_{12} = s_{11} = 0$. Therefore, every common symmetric form for these particular matrices must be singular.

Proposition 2. *If a nonzero common symmetric form S is singular, then A and B are simultaneously upper triangular in a suitable basis.*

Proof. The kernel of S has dimension one. For $v \in \ker(S)$, the identities $SA = A^T S$ and $SB = B^T S$ give $SAv = SBv = 0$. Thus $\ker(S)$ is invariant under both generators. Choosing its nonzero vector as the first basis vector makes both matrices upper triangular.

In that basis, every product X is upper triangular, with traceless part

$$X_0 = \begin{pmatrix} \alpha & \beta \\ 0 & -\alpha \end{pmatrix}.$$

The class $[\alpha : \beta]$, together with the scalar symbol $\perp$, again takes $p+2$ values, and products of the same class commute. The arguments of Section 4 therefore apply to ordinary words, without palindromes: two distinct words u_1, u_2 of length m in the same class give the collision $u_1 u_2$, $u_2 u_1$ of $2m$ bits. Together with Theorem 1, this covers every pair $A, B \in \mathrm{GL}_2(\mathbb{F}_p)$. The same commutation test, in the form $(\alpha - \alpha^{-1})\beta^{-1}$, appears in [7, Proposition 4.2], and as the code $(a - d)/b$ in the commuting-product search of [12, finding hash-04-2].

4 Commuting Palindromes

Fix A, B and an invertible symmetric S as in Proposition 1, and let

$$W = \{X \in \mathrm{M}_2(\mathbb{F}_p) \mid SX \text{ is symmetric}\}, \quad V = \{X \in W \mid \mathrm{tr}(X) = 0\}.$$

Lemma 3. *W has dimension 3 and contains I, while V has dimension 2. A word w satisfies $H_M(w) \in W$ whenever it is palindromic: $w = \mathrm{rev}(w)$.*

Proof. Multiplication by S is a linear bijection of $\mathrm{M}_2(\mathbb{F}_p)$ taking W onto the three-dimensional space of symmetric matrices. Since S is symmetric, $I \in W$. The trace is a linear form on W that is nonzero at I (as p is odd), so its kernel V has dimension 2. Finally, $X \in W$ is equivalent to $\rho_S(X) = X$, and for a palindrome $H_M(w) = H_M(\mathrm{rev}(w)) = \rho_S(H_M(w))$ by Proposition 1.

Define the *class* $\kappa(X)$ of $X \in W$ to be the line $\mathbb{F}_p X_0 \in \mathbb{P}(V)$ spanned by its traceless part when $X_0 \neq 0$, and a special symbol $\perp$ when $X_0 = 0$. There are $p + 2$ classes.

Lemma 4. *If $X, Y \in W$ have the same class, then $XY = YX$.*

Proof. If the class is $\perp$, both matrices are scalar. Otherwise $X = aI + X_0$ and $Y = bI + cX_0$ with $a, b, c \in \mathbb{F}_p$, and since both are linear in X_0, they commute.

4.1 Existence of Short Collisions

Proof (Proof of existence in Theorem 1). For $u \in \{0,1\}^m$, let $P_u = u\,\mathrm{rev}(u)$. These are 2^m distinct palindromes of equal length, and their hashes lie in W. Since $2^m \geq p + 3 > p + 2$, two distinct half-messages u_1, u_2 give hashes in the same class. Lemma 4 then provides

$$H_M(P_1P_2) = H_M(P_1)H_M(P_2) = H_M(P_2)H_M(P_1) = H_M(P_2P_1).$$

The two messages are distinct because their first $2m$ bits are the distinct words P_1 and P_2.

4.2 Evaluating the Class Map and Searching

Write $X = H_M(u)$. Proposition 1 gives

$$H_M(P_u) = X\rho_S(X) = XS^{-1}X^TS.$$

Thus a class evaluation uses m generator multiplications and a constant number of additional field operations, including one inversion to normalize the projective coordinate. If $s_{22} \neq 0$, the class of $Y \in W$ can be encoded as

$$\kappa(Y) = [Y_{11} - Y_{22} : Y_{12}],$$

with a separate scalar symbol $\perp$ when both entries vanish. Indeed, writing $S = \left(\begin{smallmatrix} s_{11} & s_{12} \\ s_{12} & s_{22} \end{smallmatrix}\right)$ and $Y_0 = \left(\begin{smallmatrix} \alpha & \beta \\ \gamma & -\alpha \end{smallmatrix}\right)$, the symmetry condition becomes $s_{22}\gamma = s_{11}\beta - 2s_{12}\alpha$, so γ is determined by α and β. For the CHAMP forms, s_{22} is 9 or 1, and is nonzero at both proposed primes.

A dictionary search samples distinct half-messages and stores one representative for each class. Under the independent-uniform-class heuristic in Theorem 1, its expected first-match cost is

$$\sqrt{\frac{\pi}{2}(p+2)}\,(1+o(1)).$$

This proves the heuristic part of Theorem 1, where uniformity of the classes was assumed. Repeated half-messages must be rejected, since they give no distinct-message collision.

For a memory-efficient search, encode the classes by $\{0,\dots,p+1\}$ and let $u(z)$ be the m-bit expansion of z. The resulting iteration is

$$f(z) = \kappa\big(H_M(u(z)\,\mathrm{rev}(u(z)))\big).$$

The encoding of z as a half-message is injective. Pollard's rho method [10], Brent's cycle detection [1], or a distinguished-point implementation [8] can be applied if this particular iteration behaves like a random function. This is actually a stronger assumption than uniformity of the classes. A recovered pair of distinct predecessors of one state yields the required collision. A walk starting on a cycle may require a restart. The $\sqrt{\pi(p+2)/2}$ constant describes the ideal first repetition, rather than the total number of evaluations used by Brent's detection and predecessor-recovery phases.

4.3 A Distribution-Independent Square-Root Bound

At half-message length m, the argument above proves that collisions exist, but the cost of finding one relies on the heuristic. Appending just one extra bit per half-message allows us to get around this heuristic.

Proposition 3. *Set $N = p+2$ and $m_+ = \lceil \log_2(2N) \rceil$, so $m_+ \le m+1$. Sample distinct $u \in \{0,1\}^{m_+}$ uniformly without replacement, and stop when two palindromes $u\,\mathrm{rev}(u)$ have the same class. If T is the number of sampled half-messages, then*

$$\Pr(T > t) \le \exp\left(-\frac{t(t-1)}{4N}\right) \quad (0 \le t \le N),$$

and $T \le N+1$. In particular, $\mathbb{E}[T] \le \sqrt{\pi N} + O(1)$. Each collision has $4m_+$ bits. No assumption on the class distribution is required.

Proof. Let $M = 2^{m_+} \ge 2N$, and let $s_1,\dots,s_N$ be the sizes of the class fibers, allowing zero sizes. For t distinct uniformly sampled inputs, the probability of no class repetition is

$$\frac{e_t(s_1,\dots,s_N)}{\binom{M}{t}},$$

where e_t is the elementary symmetric polynomial of degree t. For nonnegative s_i of sum M, pairwise averaging shows

$$e_t(s_1, \ldots, s_N) \le \binom{N}{t} \left(\frac{M}{N}\right)^t .$$

Indeed, if a, b are two of the coordinates and e'_k denotes the elementary symmetric polynomials in the remaining $N-2$, then $e_t = e'_t + (a+b)e'_{t-1} + ab\, e'_{t-2}$, with all coefficients nonnegative. For fixed $a+b$, replacing a, b by their average cannot decrease this expression. It follows that, for $t \le N$,

$$\begin{aligned} \Pr(T > t) &\le \prod_{j=0}^{t-1} \frac{1 - j/N}{1 - j/M} \\ &\le \exp\left(-\frac{t(t-1)}{2}\left(\frac{1}{N} - \frac{1}{M}\right)\right) \\ &\le \exp\left(-\frac{t(t-1)}{4N}\right). \end{aligned}$$

The middle inequality follows from $\log(1-x) - \log(1-y) \le -(x-y)$ for $0 \le y \le x < 1$. The pigeonhole principle gives $T \le N+1$. Summing the tail bound and comparing with the Gaussian integral gives $\mathbb{E}[T] \le \sqrt{\pi N} + O(1)$. Equal classes then give distinct colliding messages, exactly as in the proof of Theorem 1.

The same bound applies to the triangular class map of Proposition 2, using ordinary words and messages of $2m_+$ bits. Thus every two-generator Cayley hash into $\mathrm{GL}_2(\mathbb{F}_p)$ has an expected $O(\sqrt{p})$ dictionary attack on messages of length $O(\log(p))$. The expected field-operation cost is $O(\sqrt{p}\log(p))$, and the expected storage is $O(\sqrt{p})$ records, and neither requires a distributional heuristic.

Proof (Proof of Theorem 2). For p_{128} and p_{256}, respectively, $m = 128$ and $m = 256$. The explicit invertible forms from Section 3 give the 512-bit and 1024-bit collisions and the stated heuristic costs. For the unconditional variant, $m_+ = 129$ and $m_+ = 257$. Proposition 3 therefore gives collisions of 516 bits and 1028 bits. If byte-aligned inputs are desired, appending the same four bits to both messages gives collisions of 520 bits and 1032 bits.

Remark 2. A single collision gives a large family: all 2^t words formed by concatenating t blocks, each chosen from the distinct equal-length words P_1P_2 and P_2P_1, have the same hash. More generally, k distinct equal-length palindromes of one class give $k!$ distinct colliding words by permutation of the blocks.

5 An Explicit Small-Prime Collision

The following collision using a toy 36-bit prime was found with the search of Section 4.2, and it can be checked directly.

Set $p = 54214950071$, take the main pair, and let

$$a = 0101101110010011100011011001101001 01,$$
$$b = 0111011100011011110111011101101000 00.$$

Put $P_1 = a\,\mathrm{rev}(a)$ and $P_2 = b\,\mathrm{rev}(b)$. These are distinct 72-bit palindromes, with hexadecimal representations `5b938d9a5a59b1c9da` and `771bddda005bbbd8ee`, respectively. Direct multiplication gives

$$H_M(P_1P_2) = H_M(P_2P_1) = \begin{pmatrix} 37275322184 & 41540854233 \\ 46507229653 & 22313848993 \end{pmatrix}.$$

After entrywise inversion, the entries in CHAMP serialization order are

$$(26517283122,\ 34482230083,\ 27118424653,\ 30530760360).$$

The colliding inputs have 144 bits. We reiterate that this is a collision for the CHAMP matrices at a small toy prime, not at either of the proposed CHAMP primes. The following Julia code verifies this.

```
const p = big(54214950071)
const A = big.([-2 1; 4 -3])
const B = big.([-5 2; -6 2])
function product(s)
    X = big.([1 0; 0 1])
    for bit in s
        X = mod.(X * (bit == '0' ? A : B), p)
    end
    return X
end
a = "0101101110010011100011011001101001 01"
b = "0111011100011011110111011101101000 00"
P1 = a * reverse(a); P2 = b * reverse(b)
M1 = P1 * P2; M2 = P2 * P1
@assert M1 != M2
@assert length(M1) == length(M2) == 144
X = product(M1)
@assert X == product(M2)
inv0(x) = iszero(x) ? big(0) : invmod(x, p)
println(inv0.([X[1,1], X[2,1], X[1,2], X[2,2]]))
# [26517283122, 34482230083, 27118424653, 30530760360]
```

Listing 1.1. Julia verification code for the small-prime collision.

6 Recovery of Short Messages

For short inputs, reduction modulo p is not a lossy operation: taking centered representatives of the four entries recovers the integer product, and the message can then be read off from the integer matrix. We use the main CHAMP pair throughout this section.

For $v = (x, y)^T \in \mathbb{R}^2$ with $x \neq 0$, define its slope by $s(v) = y/x$. The matrices act on slopes via

$$s(Av) = \frac{4 - 3s(v)}{s(v) - 2}, \quad s(Bv) = \frac{2s(v) - 6}{2s(v) - 5}.$$

Lemma 5. *Define the starting interval* $I_0 = [-13/5, 3/2]$. *If* $s(v) \in I_0$, *then*

$$s(Av) \in [-59/23, 1], \quad s(Bv) \in [56/51, 3/2].$$

Both image intervals lie in I_0, *and are disjoint.*

Proof. The poles 2 and 5/2 sit outside I_0. The derivatives of the two maps are $2/(s-2)^2$ and $2/(2s-5)^2$, respectively, so both maps are increasing on I_0. Evaluating the endpoints gives the claimed intervals. The first coordinates of Av and Bv are $x(s(v)-2)$ and $x(2s(v)-5)$, which do not vanish for $s(v) \in I_0$, so the image slopes are defined.

Proof (Proof of Theorem 3). Write $H_M^{\mathbb{Z}}(X)$ for the product before reduction modulo p, and let $n = |X|$. The maximum absolute row-sum norm is submultiplicative, with $\|A\|_\infty = 7$ and $\|B\|_\infty = 8$. Hence every entry of the integer product has absolute value at most

$$8^n \le 2^{3n_0} \le \frac{p}{2}.$$

Since p is odd and the entries are integers, every entry lies strictly between $-p/2$ and $p/2$. Decoding the digest and taking centered representatives therefore recovers the exact integer matrix $M = H_M^{\mathbb{Z}}(X)$.

Let $e_1 = (1,0)^T$. Its slope is $0 \in I_0$, and Lemma 5 shows that the first-column slope of every suffix product lies in I_0. If $X = X_1 \cdots X_n$ is nonempty, then

$$Me_1 = g_{X_1} H_M^{\mathbb{Z}}(X_2 \cdots X_n) e_1.$$

Thus $s(Me_1) \le 1$ if $X_1 = 0$, and $s(Me_1) \ge 56/51$ if $X_1 = 1$. Comparing $s(Me_1)$ with 1 therefore determines the first bit. Multiplying by the corresponding inverse over $\mathbb{Q}$ recovers the integer suffix product. The inverse has denominator 2 and the divisions are exact, so iterating recovers every bit. The procedure halts at I, since a nonempty integer product has determinant $2^n \ne 1$.

Each step uses a constant number of operations on integers of $O(\log(p))$ bits. Indeed, every remaining suffix has norm at most $8^{n_0} < p/2$, and the fixed-size numerators in the inverse multiplication change this bound only by a constant factor. Digest decoding uses four modular inversions separately from these peeling steps. Finally, the first-bit rule also proves uniqueness of the recovered short word.

The slope argument proves that the integer semigroup generated by A and B is free. It does not require the sum of absolute values of the entries to decrease at every step, and it does not rely on eigenvalues of the average matrix. The deterministic bound is conservative: the same decoder works for any individual message whose integer entries all lie in $(-p/2, p/2)$. A typical-message threshold would require a specified message distribution and supporting measurements.

Short-input lifting for CHAMP has also been reported in [11], and [12, finding hash-04-3] gives a meet-in-the-middle preimage search for inputs of known

length m, of cost $O(2^{m/2})$, demonstrated on 40-bit inputs. Our contribution in this section is a decoder that runs in linear time rather than $O(2^{m/2})$, with an explicit bound covering all messages up to length n_0, and its slope-based proof. We reiterate that Theorem 3 does not apply to arbitrarily long messages, and therefore does not contradict the designers' claim, which concerns messages of at least 1000 bits.

7 Discussion

The main construction does not rely on any particular choice of A and B; indeed, any two 2×2 matrices have a nonzero common symmetric form. If the form is invertible, reversal fixes both generators and places palindrome products in a three-dimensional subspace containing I. If the form is singular, its kernel gives a common invariant line and simultaneous triangularization. In both cases, commutation reduces to equality of one of $p + 2$ classes.

Security Bounds. The distribution-independent dictionary attack gives an expected $O(\sqrt{p})$ number of class evaluations on messages of length $O(\log(p))$. For the proposed CHAMP primes, its leading exponents are 64 and 128. At the shorter lengths of 512 and 1024 bits, the cost estimate relies on the uniformity heuristic. The heuristic affects only the constant: Proposition 3 gives the $O(\sqrt{p})$ bound without it, at the cost of one extra bit per half-message. These figures are only upper bounds on the collision resistance of CHAMP, and explicit longer collisions for CHAMP-512 at the full parameters are already known [17]. The leading exponents agree with the projection in [12, finding hash-04-2]; the differences are the message length and the proven variant. For CHAMP-1024, where no explicit collision has been reported, the 2^{128} bound is the strongest collision result we know.

Design Rationale. The determinant condition constrains unequal lengths, while the palindrome attack uses equal lengths. A large girth or a large order for words hashing to I therefore does not exclude these collisions. For short messages, exact integer lifting bypasses the argument based on the sizes of modular inverses. The scalar order of a generator is another relevant invariant, and the safe-prime condition does not directly control it. The alternative pair has an additional sign-conjugacy exchanging the two generators, detailed in Appendix A.

Acknowledgments

This research was partially supported by NSF Grant DMS-2342254.

AI Disclosure

During the preparation of this manuscript, an AI assistant (Claude, Anthropic) was used to help analyze the commuting-palindrome construction, as well as

for editing purposes such as revising proofs and text. The author checked all statements and proofs, verified the example independently, and takes full responsibility for the work.

References

1. Brent, R.P.: An Improved Monte Carlo Factorization Algorithm. BIT Numerical Mathematics **20**(2), 176–184 (1980)
2. Bromberg, L., Shpilrain, V., Vdovina, A.: Navigating in the Cayley graph of $SL_2(\mathbb{F}_p)$ and Applications to Hashing. Semigroup Forum **94**(2), 314–324 (2017)
3. Charles, D.X., Lauter, K.E., Goren, E.Z.: Cryptographic Hash Functions from Expander Graphs. Journal of Cryptology **22**(1), 93–113 (2009). https://doi.org/10.1007/s00145-007-9002-x, https://doi.org/10.1007/s00145-007-9002-x
4. Demin, A., Ovchinnikov, A., Shpilrain, V.: CHAMP: Cayley HAshing with Matrix Products. Cryptology ePrint Archive, Paper 2026/2515 (2026), https://eprint.iacr.org/2026/2515
5. Grassl, M., Ilić, I., Magliveras, S., Steinwandt, R.: Cryptanalysis of the Tillich–Zémor Hash Function. Journal of Cryptology **24**(1), 148–156 (2011)
6. Idrassi, M.: CHAMP: Technical Assessment of the Submitted Specification. Technical report and reproducibility package, version 1.0.1 (Sep 2026), https://github.com/amcrypto-jp/champ-cryptanalysis, accessed 31 October 2026
7. Mullan, C., Tsaban, B.: SL_2 Homomorphic Hash Functions: Worst Case to Average Case Reduction and Short Collision Search. Cryptology ePrint Archive, Paper 2013/415 (2013), https://eprint.iacr.org/2013/415
8. van Oorschot, P.C., Wiener, M.J.: Parallel Collision Search with Cryptanalytic Applications. Journal of Cryptology **12**(1), 1–28 (1999)
9. Petit, C., Quisquater, J.J., Tillich, J.P., Zémor, G.: Hard and Easy Components of Collision Search in the Zémor–Tillich Hash Function: New Attacks and Reduced Variants with Equivalent Security. In: Topics in Cryptology – CT-RSA 2009. Lecture Notes in Computer Science, vol. 5473, pp. 182–194. Springer (2009)
10. Pollard, J.M.: Monte Carlo Methods for Index Computation (mod p). Mathematics of Computation **32**(143), 918–924 (1978)
11. PQC-X: CHAMP. PQC-X public cryptanalysis report (2026), https://www.sa-ngcc.org/c/champ, accessed 31 October 2026
12. Saarinen, M.J.O.: CHAMP (hash-04). NGCC public assessment, findings hash-04-1 to hash-04-6 (2026), https://ngcc.dev/reports/hash-04.html, accessed 31 October 2026
13. Tillich, J.P., Zémor, G.: Group-Theoretic Hash Functions. In: Algebraic Coding, First French-Israeli Workshop. Lecture Notes in Computer Science, vol. 781, pp. 90–110. Springer (1994)
14. Tillich, J.P., Zémor, G.: Hashing with SL_2. In: Advances in Cryptology – CRYPTO '94. Lecture Notes in Computer Science, vol. 839, pp. 40–49. Springer (1994). https://doi.org/10.1007/3-540-48658-5_5
15. Tillich, J.P., Zémor, G.: Collisions for the LPS Expander Graph Hash Function. In: Advances in Cryptology – EUROCRYPT 2008. Lecture Notes in Computer Science, vol. 4965, pp. 254–269. Springer (2008)
16. Voight, J.: Quaternion Algebras, Graduate Texts in Mathematics, vol. 288. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-56694-4

17. Wang, S.: Two Full CHAMP-512 Collisions. Public construction and verification package (2026), https://github.com/snipernan/champ512-collisions, accessed 31 October 2026
18. Yuan, Y., Wu, R., Wei, S., Shen, J., Liu, J., Zhang, Y.: Structural Analysis of Seven Hash Functions Submitted to the NGCC. Cryptology ePrint Archive, Paper 2026/2152 (2026), https://eprint.iacr.org/2026/2152
19. Zémor, G.: Hash Functions and Graphs with Large Girths. In: Advances in Cryptology – EUROCRYPT '91. Lecture Notes in Computer Science, vol. 547, pp. 508–511. Springer (1991)

A Additional Structure of the Alternative Pair

The alternative pair admits more symmetry than the common reversal form requires. We currently do not see how to leverage this symmetry to breach the square-root bound, but it gives further collision conditions, a second route to scalar words, and a constant-factor improvement under a uniformity assumption.

A.1 An Integer Sign-Conjugacy

Let $G = \left(\begin{smallmatrix} -1 & -1 \\ 2 & 1 \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$. Direct computation gives

$$G^2 = -I, \quad B' = GAG, \quad 3G = 5I + A - B'.$$

Since $G^{-1} = -G$, we have $GAG^{-1} = -B'$ and $GB'G^{-1} = -A$. Thus conjugation by G is an involution on the matrix algebra, although G itself has square $-I$. For the main pair, $\mathrm{tr}(A) = -5$ and $\mathrm{tr}(B) = -3$ rule out an analogous simultaneous swap by a common sign over $\mathbb{Q}$ and at the proposed primes.

Lemma 6. *For the alternative pair and every word w,*

$$H_M(\tilde{w}) = (-1)^{|w|} G H_M(w) G^{-1}.$$

Proof. Conjugation by G exchanges the generators with a minus sign, and each letter contributes one such sign.

A.2 Quaternion Coordinates

Set

$$i = G, \quad j = \begin{pmatrix} 6 & 1 \\ -10 & -6 \end{pmatrix}, \quad k = ij = \begin{pmatrix} 4 & 5 \\ 2 & -4 \end{pmatrix}.$$

Then $i^2 = -I$, $j^2 = 26I$, and $ij = -ji$. For $p \nmid 2 \cdot 3 \cdot 13$, these give a basis I, i, j, k of $\mathrm{M}_2(\mathbb{F}_p)$ and identify it with the split quaternion algebra $\left(\frac{-1,26}{\mathbb{F}_p}\right)$ [16]. Quaternion orders may also be found in the LPS and Pizer hash functions [3], but there the quaternion algebra is definite, whereas here it is the split algebra $\mathrm{M}_2(\mathbb{F}_p)$ itself, and the coordinates only give the symmetries for the alternative pair. Quaternion conjugation is the adjugate:

$$\mathrm{adj}(x_0 I + x_1 i + x_2 j + x_3 k) = x_0 I - x_1 i - x_2 j - x_3 k.$$

The generators are

$$A = \frac{1}{2}(-5I + 3i + k), \quad B' = \frac{1}{2}(5I - 3i + k).$$

They generate the full matrix algebra, since $k = A + B'$, $i = (A - B' + 5I)/3$, and $ik = -j$.

Proposition 4. *Write $H_M(w) = x_0 I + x_1 i + x_2 j + x_3 k$ and $n = |w|$ for the alternative pair. Then*

$$H_M(\mathrm{rev}(w)) = x_0 I + x_1 i - x_2 j + x_3 k, \tag{1}$$

$$H_M(\tilde{w}) = (-1)^n (x_0 I + x_1 i - x_2 j - x_3 k), \tag{2}$$

$$H_M(\mathrm{rev}(\tilde{w})) = (-1)^n (x_0 I + x_1 i + x_2 j - x_3 k). \tag{3}$$

Proof. Conjugation by i fixes I, i and negates j, k, proving (2) by Lemma 6. The anti-automorphism $\sigma(X) = j\,\mathrm{adj}(X)j^{-1}$ fixes I, i, k and negates j. Both σ and $\rho_{\widehat{S}}$ are anti-automorphisms fixing A and B', so they agree on the algebra these generate – which is all of $\mathrm{M}_2(\mathbb{F}_p)$. Proposition 1 gives (1), and composing the two identities gives (3).

Corollary 1. *For $X = H_M(w)$ and $n = |w|$, the following hold.*

1. *If $x_0 = 0$, then $H_M(ww)$ is scalar.*
2. *If $x_1 = 0$, then $H_M(w\tilde{w}) = (-1)^n 2^n I$.*
3. *If $x_2 = 0$ and $w \neq \mathrm{rev}(w)$, then w and $\mathrm{rev}(w)$ are a collision.*
4. *If $x_3 = 0$, n is even, and $w \neq \mathrm{rev}(\tilde{w})$, then w and $\mathrm{rev}(\tilde{w})$ are a collision.*

Proof. The first claim is Cayley–Hamilton with $\mathrm{tr}(X) = 2x_0 = 0$. For the second, $x_1 = 0$ implies $H_M(\tilde{w}) = (-1)^n \mathrm{adj}(X)$, and $X\,\mathrm{adj}(X) = 2^n I$. The remaining claims follow from Proposition 4.

The second item gives a method of finding scalar words that is independent of the trace condition. Since $x_1 = -\mathrm{tr}(XG)/2$, it can also be written as $\mathrm{tr}(H_M(w)G) = 0$.

A.3 Two Structured Word Families

Palindromes have $x_2 = 0$. Words of the form $u\,\mathrm{rev}(\tilde{u})$ have even length and satisfy $w = \mathrm{rev}(\tilde{w})$, hence $x_3 = 0$. Each family therefore lies in a three-dimensional subspace containing I, and proportional traceless parts again imply commutation.

For ordinary palindromes, letter exchange acts on the projective class by

$$[x_1 : x_3] \mapsto [x_1 : -x_3].$$

There are two fixed projective points, so the projective line has $(p+3)/2$ orbits under this involution. Including the scalar class gives $(p+5)/2$ orbits in total. If

two palindromes P, Q lie in the same orbit, then P has the class of Q or of $\tilde{Q}$, and the corresponding products commute. To avoid the trivial match $P = \tilde{Q}$, sample only half-messages whose first bit is 0. Every sampled palindrome then begins with 0 and every complement with 1, so we get for free that the two concatenated messages are distinct. Under an approximately uniform orbit model, this reduces the birthday target by about a factor of 2 and the sample count by about $\sqrt{2}$.

B Scalar Words and Generator Orders

Lemma 7. *If $H_M(Y)$ is scalar, then $H_M(XY) = H_M(YX)$ for every word X. This is a collision only when $XY \neq YX$. If $\mathrm{tr}(H_M(w)) = 0$, then $H_M(ww)$ is scalar.*

Proof. A scalar matrix commutes with every matrix. For a trace-zero matrix X, Cayley–Hamilton gives $X^2 = -\det(X)I$.

A trace-zero search would take about p trials if the tested word products had approximately uniform trace. The linear condition alone does not establish that heuristic or even ensure that the sampled family reaches trace zero. The main class attack avoids relying on a scalar word.

For a generator g with distinct eigenvalues λ_1, λ_2, the power g^e is scalar exactly when $(\lambda_1/\lambda_2)^e = 1$. Its scalar order $r(g)$ is therefore the multiplicative order of this ratio. The characteristic polynomial of A is $x^2 + 5x + 2$, with discriminant 17, while B has eigenvalues $-1, -2$.

Remark 3. For both CHAMP primes, $r(B) = q$ and $B^q = -I$. For p_{256}, 17 is a square and direct modular exponentiation gives $A^q = I$. Since q is prime and A is not scalar, $r(A) = q$. For p_{128}, 17 is a nonsquare and the eigenvalue ratio lies in the norm-one subgroup of order $p+1$. The complete factorization relevant to the scalar order is

$$p_{128} + 1 = 2^3 \cdot 3^3 \cdot 903785901572855597 \cdot 1743091284778891979.$$

The two final factors are prime. Writing $r_0 = (p_{128} + 1)/8$, modular exponentiation shows that A^{r_0} is scalar while $A^{r_0/\ell}$ is not, for each prime $\ell \in \{3, 903785901572855597, 1743091284778891979\}$. Hence $r(A)$ divides r_0 but no r_0/ℓ, so $r(A) = r_0 \approx 2^{125}$. At the smaller prime $p = 1012463$, the same test gives $r(A) = 237$, which is much smaller. With $r = r(A)$, the scalar power A^r gives collisions between $0^r w$ and $w0^r$ whenever these words are distinct. The safe-prime condition on $p - 1$ does not control the factorization of $p + 1$ or, by itself, scalar orders in the nonsplit case.