hexagon:2610.00301v2

The Exact Junta Bound for Degree-Two Boolean Functions on q-ary Hamming Cubes

Yuran Xiu

14 October 2026

Abstract

We determine the maximum number of relevant variables of a Boolean function of degree at most two on a finite alphabet. For every integer $q \geq 2$, this maximum is exactly $\max\{4, q+1\}$. Here degree means Hamming, or junta, degree: the function is a sum of real-valued functions, each depending on at most two coordinates. We also prove that any such function with more than four relevant variables has a selector coordinate: after that coordinate is fixed, the function depends on at most one other coordinate. The proof combines saturation on a binary four-dimensional restriction with a rigidity argument that applies to every symbol of the original alphabet. Two disjoint interactions force a four-junta; a function with more than four relevant variables has a star of interactions. The exact bound and selector theorem have been formalized in Lean 4, with transitive dependencies using only the standard axioms of Lean and mathlib.

1 Introduction

A Boolean function on a product space may be given as a sum of terms that each involve only a few coordinates. How many coordinates can affect its output? This is the junta problem for bounded-degree Boolean functions. Write $R_{d,q}$ for the largest possible number of relevant coordinates of a Boolean function on $[q]^n$, over all dimensions n, whose Hamming degree is at most d. We use $[q] = \{1, \ldots, q\}$ and give a definition of degree that does not require an algebraic structure on the alphabet.

Definition 1.1. A real-valued function has *degree at most d* if it is a finite sum of real-valued functions, each depending on at most d coordinates. A coordinate i is *relevant* for f if there exist inputs x, y differing only at i with $f(x) \neq f(y)$. Let $\mathrm{Rel}(f)$ be the set of relevant coordinates. A function with at most r relevant coordinates is an r-junta.

For a cyclic alphabet, this degree is the maximum support size of a nonzero Fourier character. It counts the number of coordinates in an interaction, rather than the ordinary total exponent after symbols are represented by integers. In particular, every unary function has degree at most one. Constants and linear terms are permitted throughout. The Fourier equivalence follows directly from orthogonality: a function depending on a set S has no Fourier coefficient on a character involving a coordinate outside S, and a character depends only on its support. For a real-valued function, taking real parts of the Fourier expansion gives real-valued summands with the same support bounds. This is also the finite-sum definition used in [4, Definition 2].

Theorem 1.2 (Exact junta bound)**.** *For every integer $q \geq 2$, every $n \geq 0$, and every Boolean function $f : [q]^n \to \{0, 1\}$ of degree at most two,*

$$|\,\mathrm{Rel}(f)| \leq \max\{4, q+1\}.$$

For every $q \geq 2$, equality is attained by a function on exactly $\max\{4, q+1\}$ *coordinates. Thus*

$$R_{2,q} = \max\{4, q+1\}. \tag{1}$$

The upper bound follows from a structural alternative.

Theorem 1.3 (Selector structure)**.** *Let A be a finite nonempty alphabet, and let $f : A^n \to \{0,1\}$ have degree at most two. If $|\operatorname{Rel}(f)| > 4$, there exists a coordinate i such that, for each $a \in A$, the section $f|_{x_i=a}$ depends on at most one coordinate other than i.*

The remaining coordinate may depend on a, and a section may be constant. Consequently such a function can be evaluated by first querying the selector and then, if needed, querying one additional coordinate. The statement does not assert a two-query algorithm for the four-junta alternative.

Prior work. The binary case $R_{2,2} = 4$ is classical. Filmus gives an explicit classification of quadratic Boolean functions on the binary cube [1]. Mogilnykh and Valyuzhenich classify equitable two-partitions of Hamming graphs associated with the second eigenvalue [2]. For Boolean functions, their setting restricts the spectrum to levels 0 and 2. The present theorem allows arbitrary level-one terms, and makes no equitability or balance assumption.

For general degree and $q \geq 3$, Valyuzhenich proved $R_{d,q} \leq dq^{d+1}/(4(q-1))$ [3]. More recently, he proved

$$R_{d,q} \leq m_q q^d, \qquad m_q = \frac{2q(q^2+4q+1)}{(q-1)^4}, \tag{2}$$

with improved constants for several small alphabets [4]. At degree two, (2) is $2q + O(1)$ as $q \to \infty$. The addressor construction giving $R_{2,q} \geq q+1$ is also recorded in [4]. Our contribution is the exact unrestricted degree-two bound and the selector alternative. The binary saturation argument and the matching constructions are known ingredients.

Proof overview. The pair interactions of a quadratic function define a graph. If that graph has two disjoint edges, their witnesses can be placed on one binary four-dimensional restriction. A Boolean quadratic function on four relevant binary variables saturates the influence bound. We use this saturation to show that every coordinate outside the restriction is irrelevant on the entire alphabet. If the graph has no disjoint edges, an elementary graph argument reduces it to a star or a triangle. Boolean range rules out independent nonconstant components, and the triangle alternative involves only three coordinates. Fixing the center of the star leaves an additive Boolean function, hence a one-junta.

2 Anchored expansions and interactions

The arguments in this section apply to a product $\Omega = \prod_{i=1}^n A_i$ of finite nonempty alphabets. Choose an anchor $o \in \Omega$. For $x \in \Omega$ and $S \subseteq [n]$, write $o[S \leftarrow x]$ for the tuple agreeing with x on S and with o elsewhere. We also use $x[i \leftarrow a]$ for a one-coordinate replacement.

Lemma 2.1 (Anchored expansion)**.** *If $f : \Omega \to \mathbb{R}$ has degree at most two, then*

$$f(x) = c + \sum_i u_i(x_i) + \sum_{i<j} v_{ij}(x_i, x_j), \tag{3}$$

where

$$\begin{aligned} c &= f(o), \\ u_i(a) &= f(o[i \leftarrow a]) - f(o), \\ v_{ij}(a,b) &= f(o[i \leftarrow a, j \leftarrow b]) - f(o[i \leftarrow a]) - f(o[j \leftarrow b]) + f(o). \end{aligned}$$

The function u_i vanishes at o_i, and v_{ij} vanishes when either argument is its anchor symbol. Conversely, every function of the form (3) has degree at most two.

Proof. All operators in the displayed formula are linear in f. It therefore suffices to check a summand depending on zero, one, or two coordinates. For a constant only c remains. For a unary summand, c and its unary correction reconstruct it. For a summand on i, j, these terms together with v_{ij} reconstruct the summand, and every other correction is zero. The converse follows directly from Definition 1.1. □

For distinct i, j, symbols $a, a' \in A_i$, $b, b' \in A_j$, and a fixed background in the other coordinates, define the mixed difference

$$\begin{aligned} \Delta_{ij} f(a, a'; b, b') = f(x[i \leftarrow a, j \leftarrow b]) &- f(x[i \leftarrow a', j \leftarrow b]) \\ &- f(x[i \leftarrow a, j \leftarrow b']) + f(x[i \leftarrow a', j \leftarrow b']). \end{aligned}$$

For a degree-two function this quantity is independent of the background: in (3), all terms except v_{ij} cancel. Likewise, a difference of two sections in one coordinate has degree at most one on the remaining coordinates. Restrictions to subalphabets preserve degree.

Definition 2.2. The *interaction graph* of f has vertex set $[n]$ and an edge ij when some mixed difference $\Delta_{ij} f$ is nonzero. Equivalently, v_{ij} is not identically zero. This graph is independent of the anchor.

An interaction makes both of its endpoints relevant. Restricting to a subalphabet can remove interactions or relevance. In the proof below, every binary restriction is chosen to preserve specified nonzero witnesses.

Lemma 2.3 (Independent additive parts)**.** *Suppose a Boolean function is a sum of real-valued functions on disjoint sets of coordinates. At most one summand is nonconstant. In particular, a Boolean function of degree at most one depends on at most one coordinate.*

Proof. If two summands have values $a < a'$ and $b < b'$, respectively, independence of their coordinates realizes, with all other summands fixed, the three strictly ordered sums $a + b < a' + b < a' + b'$. A Boolean function has only two possible values. The last assertion follows from the unary version of (3). □

Grouping (3) by connected components of the interaction graph gives independent additive parts. A component containing an edge has a nonconstant part, because its corresponding mixed difference is nonzero. An isolated vertex is relevant precisely when its unary part is nonconstant. Lemma 2.3 consequently implies the following.

Corollary 2.4. *For a nonconstant Boolean function of degree at most two, all relevant coordinates lie in a single connected component of its interaction graph. An isolated relevant coordinate forces the entire function to be unary.*

3 Binary saturation and full-alphabet rigidity

We first recall the binary equality case needed for the argument. On $\{-1,1\}^4$, use the uniform measure and the Fourier characters $\chi_S(z) = \prod_{i\in S} z_i$. For a real-valued function g, write $\widehat{g}(S) = \mathbb{E}[g\chi_S]$.

Lemma 3.1 (Binary saturation)**.** *Let $g : \{-1,1\}^4 \to \{-1,1\}$ have degree at most two and depend on all four coordinates. Then*

$$g(z) = \sum_{|S|=2} \widehat{g}(S)\chi_S(z), \qquad \sum_{|S|=2} \widehat{g}(S)^2 = 1. \tag{4}$$

For every $i \in [4]$,

$$\sum_{\substack{|S|=2\\ i\notin S}} \widehat{g}(S)^2 = \frac{1}{2}. \tag{5}$$

In particular, among the other three coordinates there is a nonzero quadratic coefficient.

Proof. For each i, the half-difference

$$D_i g = \frac{g|_{z_i=1} - g|_{z_i=-1}}{2}$$

is a nonzero affine function of the other three bits and takes values in $\{-1,0,1\}$. A nonzero affine function on a binary cube is nonzero on at least half the inputs. Indeed, if a variable coefficient is nonzero, at most one of the two points in each fiber of that variable is a zero; if all variable coefficients vanish, the function is a nonzero constant. It follows that $\mathrm{Inf}_i(g) := \mathbb{E}[(D_i g)^2] \geq 1/2$.

Parseval's identity and the degree bound yield

$$2 \leq \sum_{i=1}^{4} \mathrm{Inf}_i(g) = \sum_S |S|\widehat{g}(S)^2 \leq 2\sum_S \widehat{g}(S)^2 = 2.$$

Equality throughout forces the constant and linear coefficients to vanish, and gives $\mathrm{Inf}_i(g) = 1/2$ for every i. Subtracting the quadratic mass containing i from the total mass one proves (5). □

Lemma 3.2 (A binary restriction determines all outside coordinates)**.** *Let $F : \prod_{i=1}^n A_i \to \{-1,1\}$ have degree at most two. Suppose there is a set I of four coordinates, two distinct chosen symbols in each coordinate of I, and fixed values for the remaining coordinates, such that the resulting binary restriction depends on all four coordinates. Then F depends only on the coordinates in I on the full product.*

Proof. Identify each chosen symbol pair with $\{-1,1\}$. For every assignment t to the coordinates outside I, let g_t be the resulting function on the same four-dimensional binary cube. Let $g = g_{t_0}$ be the restriction in the hypothesis. Every quadratic Fourier coefficient of g_t is the same as that of g: mixed differences inside I are independent of all outside coordinates. By Lemma 3.1, these shared quadratic coefficients already have squared norm one. Since g_t is sign-valued, Parseval forces all of its remaining coefficients to vanish. Thus

$$g_t = g \qquad \text{for every outside assignment } t. \tag{6}$$

It remains to handle symbols of the coordinates in I that are absent from the chosen binary cube. Fix $j \notin I$ and any $a, b \in A_j$, and set

$$D(x) = \frac{F(x[j \leftarrow a]) - F(x[j \leftarrow b])}{2},$$

regarding D as a function of all coordinates other than j. This function has degree at most one and takes values in $\{-1, 0, 1\}$. By (6), it vanishes whenever the coordinates in I lie in their chosen symbol pairs, for every assignment of the other coordinates.

Choose an anchor with chosen symbols on I. The unary anchored expansion of D therefore has zero constant term and zero unary terms outside I. It has the form

$$D(x) = \sum_{i \in I} r_i(x_i), \tag{7}$$

where r_i vanishes on both chosen symbols of coordinate i. To see the vanishing of each term, vary only that coordinate from the anchor and use the vanishing on the entire cylinder.

Fix $i \in I$ and an arbitrary symbol $s \in A_i$. On the slice $x_i = s$ with the other three coordinates of I restricted to their chosen pairs, equation (7) gives $D = r_i(s)$. If $r_i(s) \neq 0$, then it is 1 or -1. Two signs with half-difference 1, or with half-difference -1, are forced to be the corresponding opposite constants. Hence both $F|_{x_j=a,x_i=s}$ and $F|_{x_j=b,x_i=s}$ are constant on this three-cube.

On the other hand, (5) gives a nonzero quadratic coefficient of g involving two coordinates in $I \setminus \{i\}$. Its mixed difference is unchanged when x_i is replaced by s or x_j by a or b. It is therefore nonzero in the supposedly constant slice, a contradiction. This proves $r_i(s) = 0$ for every $i \in I$ and every symbol s. Equation (7) now gives $D \equiv 0$ on the full product.

The coordinate j and its symbols a, b were arbitrary. Every coordinate outside I is consequently irrelevant. Changing such coordinates one at a time shows that F depends only on I. □

The use of the global unary expansion in (7) is essential: it handles assignments in which several coordinates of I simultaneously take new symbols. Merely proving invariance on one selected binary cube would not establish the conclusion.

Corollary 3.3 (Disjoint interactions force four variables)**.** *If the interaction graph of a Boolean degree-two function contains two disjoint edges ij and $k\ell$, the function depends only on $\{i, j, k, \ell\}$.*

Proof. Choose two symbols in each endpoint of ij witnessing its nonzero mixed difference, and do the same for $k\ell$. These choices involve disjoint coordinates, so they define a single binary four-cube. Both interactions remain nonzero on this cube, regardless of the outside assignment. Thus all four coordinates are relevant there. Apply Lemma 3.2 to $F = 2f - 1$. □

4 The selector alternative and the upper bound

Lemma 4.1 (Intersecting edges)**.** *A simple graph in which every two edges intersect either has a vertex incident with every edge, or has all of its edges supported on three vertices. The empty graph is included in the second alternative.*

Proof. With at most one edge the assertion is immediate. Otherwise take distinct edges ab and ac. If every edge contains a, the first alternative holds. An edge avoiding a must be bc. Any edge meeting all three of ab, ac, bc has both endpoints in $\{a, b, c\}$, proving the second alternative. □

Proof of Theorem 1.3. Suppose $|\operatorname{Rel}(f)| > 4$. By Corollary 3.3, every two interaction edges intersect. The graph cannot be empty, since otherwise f is additive and Lemma 2.3 makes it a one-junta. By Corollary 2.4, all relevant coordinates belong to its single nonconstant component. The alternative in which all edges are supported on three vertices would therefore give at most three relevant coordinates. Lemma 4.1 consequently provides a common endpoint i of all edges.

Fix any $a \in A$ at coordinate i. In (3), each remaining pair term contains i, so after this substitution it becomes unary. The section is an additive Boolean function. Lemma 2.3 shows that it depends on at most one of the remaining coordinates. □

Upper bound in Theorem 1.2. There is nothing to prove when $|\operatorname{Rel}(f)| \le 4$. Otherwise take the selector i given by Theorem 1.3. For each $a \in [q]$, let $S_a \subseteq [n] \setminus \{i\}$ be a set of size at most one supporting the section $f|_{x_i=a}$.

If a coordinate $j \neq i$ is relevant to f, its two witnessing inputs have the same value a in coordinate i. It is then relevant to that section, so $j \in S_a$. Therefore

$$\operatorname{Rel}(f) \subseteq \{i\} \cup \bigcup_{a\in[q]} S_a, \qquad |\operatorname{Rel}(f)| \le 1 + \sum_{a\in[q]} |S_a| \le q+1.$$

This also covers all small dimensions through the initial four-variable alternative; constants and the zero-dimensional cube have no relevant coordinates. □

5 Matching examples

We include the known examples to establish attainment for every alphabet.

The addressor. Choose a nonconstant $b : [q] \to \{0,1\}$. On $q+1$ coordinates define

$$f(s, y_1, \dots, y_q) = b(y_s) = \sum_{a=1}^{q} \mathbf{1}_{\{s=a\}} b(y_a). \tag{8}$$

Each summand depends on only s and y_a, so the degree is at most two. Every y_a is relevant by setting $s = a$. The selector is relevant because $q \ge 2$ and two leaf coordinates can be assigned different b-values. Hence $R_{2,q} \ge q+1$. This is the degree-two addressor underlying the general decision-tree lower construction in [4].

The binary square. For signs a, b, c, d, the standard quadratic example

$$G(a,b,c,d) = \frac{(a+b)c + (a-b)d}{2} \tag{9}$$

is always sign-valued: if $a = b$ it equals ac, and if $a = -b$ it equals ad. It is quadratic and all four coordinates are relevant. Indeed, c matters when $a = b$, d matters when $a = -b$, a matters when $c = d$, and b matters when $c = -d$. This is a form of the familiar four-variable quadratic Boolean example; see [1].

For any $q \ge 2$, choose a surjection $\sigma : [q] \to \{-1, 1\}$ and define

$$h(x_1, x_2, x_3, x_4) = \frac{G(\sigma(x_1), \sigma(x_2), \sigma(x_3), \sigma(x_4)) + 1}{2}.$$

The function h is Boolean, has degree at most two, and depends on all four coordinates. No restriction on the sizes of the fibers of σ is needed. Thus $R_{2,q} \ge 4$. Together with (8) and the upper bound, this proves Theorem 1.2.

6 Formal verification

The exact bound and selector theorem have been formalized in Lean 4. The formal degree hypothesis is literally the finite-sum formulation in Definition 1.1: the summands are real-valued and each has at most two globally relevant coordinates. Boolean range is required of f, not of the summands. The formal result includes arbitrary natural dimensions, all integers $q \geq 2$, and the matching lower examples.

The main declarations, in namespace `QaryQuadratic`, are

Declaration	Statement
`sharp_qary_quadratic`	Theorem 1.2, including attainment
`source_quadratic_selector`	Theorem 1.3

The formal proof uses the same interaction and rigidity mechanism as the mathematical proof. For the binary step, it proves a finite saturation lemma for a four-cube with two disjoint nonzero interactions. The eleven Boolean entries at inputs of Hamming weight at most two determine the remaining five entries by the degree-two identities. Ordinary kernel reduction verifies the required evenness and balance conditions. The interface to this finite lemma derives its hypotheses from an arbitrary real degree-two Boolean function. It does not assume a table constraint in the main theorem. This supplies the uniqueness consequence of binary saturation used in the cylinder argument.

The rest of the development proves the passage to arbitrary alphabets and dimensions, including arbitrary outside backgrounds, arbitrary new symbols in the four selected coordinates, the interaction-graph alternative, selector counting, and coordinate relabeling for the lower examples. The formal graph argument uses the isolated-coordinate case of Corollary 2.4 in place of an explicit connected-component decomposition. A separate formal equivalence with complex Fourier degree is not claimed; the checked statement uses the finite-sum definition.

Versions and trust boundary. The verification uses Lean 4.27.0, with compiler commit

`db93fe1608548721853390a10cd40580fe7d22ae`,

and mathlib commit

`a3a10db0e9d66acbebf76c5e6a135066525ac900`.

The source package submitted to arXiv contains the Lean sources and verification records under `anc/lean/`; the Hexagon deposit contains the manuscript only. The artifact records the dependency manifest, source hashes, reproduction scripts, successful full-build output, and transitive axiom reports. An independent rebuild compiled all twelve local proof modules and the umbrella module from source, excluding earlier local compiled artifacts. The two main declarations depend only on `propext`, `Classical.choice`, and `Quot.sound`. There are no admitted lemmas or additional mathematical axioms. The finite auxiliary lemma uses kernel `decide`, with no `native_decide` or external numerical certificate. These checks concern the formal statements and proof objects; the mathematical exposition above is a separate presentation of the argument.

References

[1] Yuval Filmus. Quadratic Boolean functions on the cube. Manuscript, 27 March 2022. https://yuvalfilmus.cs.technion.ac.il/Manuscripts/Degree2.pdf.

[2] Ivan Mogilnykh and Alexandr Valyuzhenich. Equitable 2-partitions of the Hamming graphs with the second eigenvalue. *Discrete Mathematics*, 343(11):112039, 2020. doi:10.1016/j.disc.2020.112039. Preprint: arXiv:1903.12333.

[3] Alexandr Valyuzhenich. An upper bound on the number of relevant variables for Boolean functions on the Hamming graph. *Discrete Mathematics*, 349(2):114745, 2026. doi:10.1016/j.disc.2025.114745. Preprint: arXiv:2404.10418v1, 2024.

[4] Alexandr Valyuzhenich. An upper bound on the number of relevant variables in a bounded degree Boolean function on the Hamming graph. arXiv:2609.12540v1, 2026.