Linearity Matters: Genericity and Existence Results for Linear and Nonlinear Codes

Anna-Lena Horlemann

University of St.Gallen, Switzerland

Encode - Eindhoven
June 23rd 2026



The naive expectation

all codes	$\supseteq$	linear codes
$C \subseteq \mathbb{F}_q^n$		$C \leq \mathbb{F}_q^n$
$\binom{q^n}{q^k} \approx q^{(n-k)q^k}$	$\gg$	$\begin{bmatrix} n \\ k \end{bmatrix}_q \approx q^{(n-k)k}$
many more choices (for good codes)		less choices

The naive expectation

all codes	$\supseteq$	linear codes
$C \subseteq \mathbb{F}_q^n$		$C \leq \mathbb{F}_q^n$
$\binom{q^n}{q^k} \approx q^{(n-k)q^k}$	$\gg$	$\begin{bmatrix} n \\ k \end{bmatrix}_q \approx q^{(n-k)k}$
many more choices (for good codes)		less choices

BUT...

1 The Naive Expectation

2 Known Bounds

3 Behavior of Random Codes (Genericity)

4 Linearity as a Constraint

5 The Big Picture

Upper bounds on cardinality I

possible metrics: Hamming, rank, sum-rank

Most upper bounds do not impose any linearity constraints.

- **Singleton bound**[1]

$$\log_q \underbrace{A_q(n,d)}_{S} \leq n - d + 1$$

$$\log_q \underbrace{B_q(n,d)}_{q^k} = k \leq n - d + 1$$

$\Longrightarrow$ Exactly the same for linear and non-linear codes

[1] Parameters need to be chosen correctly in the various metric spaces.

Upper bounds on cardinality II

possible metrics: Hamming, rank, sum-rank, possibly Lee

Most upper bounds do not impose any linearity constraints.

- **Sphere packing bound**

$$\underbrace{A_q(n,d)}_{S} \leq \frac{q^n}{V_q(n, \lfloor \frac{d-1}{2} \rfloor)} \iff \log_q S \leq n - \log_q V_q(n, \lfloor \frac{d-1}{2} \rfloor)$$

$$\underbrace{B_q(n,d)}_{q^k} \leq \frac{q^n}{V_q(n, \lfloor \frac{d-1}{2} \rfloor)} \iff k \leq n - \lceil \log_q V_q(n, \lfloor \frac{d-1}{2} \rfloor) \rceil$$

Upper bounds on cardinality II

possible metrics: Hamming, rank, sum-rank, possibly Lee

Most upper bounds do not impose any linearity constraints.

- **Sphere packing bound**

$$\underbrace{A_q(n,d)}_{S} \leq \frac{q^n}{V_q(n, \lfloor \frac{d-1}{2} \rfloor)} \iff \log_q S \leq n - \log_q V_q(n, \lfloor \frac{d-1}{2} \rfloor)$$

$$\underbrace{B_q(n,d)}_{q^k} \leq \frac{q^n}{V_q(n, \lfloor \frac{d-1}{2} \rfloor)} \iff k \leq n - \lceil \log_q V_q(n, \lfloor \frac{d-1}{2} \rfloor) \rceil$$

$\implies$ The only difference for linear or non-linear codes comes from (logarithmic) rounding.

$\implies$ Differ at most in a factor of q (asymptotically in n negligible; in q it disappears for our metrics).

Upper bound for linear codes

Hamming metric

- **Griesmer bound**

$$\sum_{i=0}^{k-1} \lceil \frac{d}{q^i} \rceil \leq n$$

Usually stronger than the Singleton bound. Proof assumes linearity, i.e., it does not directly apply to non-linear codes.

Lower bounds (Gilbert-Varshamov)

possible metrics: Hamming, rank, sum-rank, possibly Lee

- **Gilbert-Varshamov bound**

$$\underbrace{A_q(n,d)}_{S} \geq \frac{q^n}{V_q(n,d-1)} \iff \log_q S \geq n - \log_q V_q(n,d-1)$$

$$\underbrace{B_q(n,d)}_{q^k} > \frac{q^n}{V_q(n,d-1)} \iff k > n - \lfloor \log_q V_q(n,d-1) \rfloor$$

Lower bounds (Gilbert-Varshamov)

possible metrics: Hamming, rank, sum-rank, possibly Lee

- **Gilbert-Varshamov bound**

$$\underbrace{A_q(n,d)}_{S} \geq \frac{q^n}{V_q(n,d-1)} \iff \log_q S \geq n - \log_q V_q(n,d-1)$$

$$\underbrace{B_q(n,d)}_{q^k} > \frac{q^n}{V_q(n,d-1)} \iff k > n - \lfloor \log_q V_q(n,d-1) \rfloor$$

$\Longrightarrow$ Although statements are almost the same, the proofs for linear and non-linear codes are very different.

$\Longrightarrow$ The bounds differ at most in a factor of q.

So...

The bounds do not differ (a lot) for linear and non-linear codes!

(Except maybe the Griesmer bound for linear Hamming metric codes.)

1 The Naive Expectation

2 Known Bounds

3 Behavior of Random Codes (Genericity)

4 Linearity as a Constraint

5 The Big Picture

Asymptotic density of non-linear codes

$$a_n \in o(f_n) \iff \lim a_n/f_n = 0 \quad , \quad a_n \in \omega(f_n) \iff \lim f_n/a_n = 0$$

Theorem

$$\lim \delta_q^D(\mathbb{F}_q^n, S_{[n,q,d]}, 0, d) = \begin{cases} 1 & \text{if } \mathbf{v}_q^D(\mathbb{F}_q^n, d-1) \in o(q^n S_{[n,q,d]}^{-2}) \\ 0 & \text{if } \mathbf{v}_q^D(\mathbb{F}_q^n, d-1) \in \omega(q^n S_{[n,q,d]}^{-2}) \end{cases}$$

as q *or* $n \to +\infty$.

Asymptotic density of non-linear codes

$$a_n \in o(f_n) \iff \lim a_n/f_n = 0 \quad , \quad a_n \in \omega(f_n) \iff \lim f_n/a_n = 0$$

Theorem

$$\lim \delta_q^D(\mathbb{F}_q^n, S_{[n,q,d]}, 0, d) = \begin{cases} 1 & \text{if } \mathbf{v}_q^D(\mathbb{F}_q^n, d-1) \in o(q^n S_{[n,q,d]}^{-2}) \\ 0 & \text{if } \mathbf{v}_q^D(\mathbb{F}_q^n, d-1) \in \omega(q^n S_{[n,q,d]}^{-2}) \end{cases}$$

as q or $n \to +\infty$.

GV-bound:

$$S_{[n,q,d]} \geq \frac{q^n}{\mathbf{v}_q^D(\mathbb{F}_q^n, d-1)}$$

Corollary

Non-linear codes achieving the Gilbert-Varshamov bound are asymptotically sparse with respect to q or n.

Asymptotic density of (sub-)linear codes

For q^s-linearity: $m = s\ell$

Theorem

$$\lim \delta_q^D(\mathbb{F}_{q^m}^n, q^{\ell k}, \ell, d) = \begin{cases} 1 & \text{if } \mathbf{v}_q^D(\mathbb{F}_{q^m}^n, d-1) \in o(q^{\ell(ns+1-k)}), \\ 0 & \text{if } \mathbf{v}_q^D(\mathbb{F}_{q^m}^n, d-1) \in \omega(q^{\ell(ns+1-k)}), \end{cases}$$

as q, n, s *or* $\ell \to +\infty$.

Asymptotic density of (sub-)linear codes

For q^s-linearity: $m = s\ell$

Theorem

$$\lim \delta_q^D(\mathbb{F}_{q^m}^n, q^{\ell k}, \ell, d) = \begin{cases} 1 & \text{if } \mathbf{v}_q^D(\mathbb{F}_{q^m}^n, d-1) \in o(q^{\ell(ns+1-k)}), \\ 0 & \text{if } \mathbf{v}_q^D(\mathbb{F}_{q^m}^n, d-1) \in \omega(q^{\ell(ns+1-k)}), \end{cases}$$

as q, n, s *or* $\ell \to +\infty$.

Corollary

1. *(Sub-)linear codes achieving the Gilbert-Varshamov bound are asymptotically dense with respect to* q *or* ℓ.
2. *The asymptotic density of (sub-)linear codes achieving the Gilbert-Varshamov bound is upper bounded by* $q^\ell/(q^\ell + 1)$, *with respect to* n *or* s.

Asymptotics for fixed rate

$$R_{GV} := 1 - \frac{1}{n} \log_{q^\ell} \mathbf{v}_q^D(\mathbb{F}_{q^m}^n, d-1)$$

Corollary

The density of (sub-)linear codes achieving the Gilbert-Varshamov bound is (asymptotically in n, or not) bounded by

$$\frac{q^\ell - 1}{q^\ell} \leq \delta_q^D(\mathbb{F}_{q^m}^n, q^{\ell R_{GV} n}, \ell, d) \leq \frac{q^\ell}{q^\ell + 1}.$$

Asymptotics for fixed rate

$$R_{GV} := 1 - \frac{1}{n} \log_{q^\ell} \mathbf{v}_q^D(\mathbb{F}_{q^m}^n, d-1)$$

Corollary

The density of (sub-)linear codes achieving the Gilbert-Varshamov bound is (asymptotically in n, or not) bounded by

$$\frac{q^\ell - 1}{q^\ell} \leq \delta_q^D(\mathbb{F}_{q^m}^n, q^{\ell R_{GV} n}, \ell, d) \leq \frac{q^\ell}{q^\ell + 1}.$$

Corollary

(Sub-)linear codes achieving the Gilbert-Varshamov bound are asymptotically dense with respect to q or ℓ.

Asymptotic Bounds for GV-Codes in n

BUT :

Theorem

(Sub-)linear codes with a rate arbitrarily close to the GV-bound are also dense with respect to n, s:

$$\lim_{n,s\to\infty} \delta_q^D(\mathbb{F}_{q^m}^n, q^{\ell(R_{GV}-\epsilon)n}, \ell, d) = 1.$$

Hamming Metric Asymptotics

$$\mathbf{v}_q^{\mathrm{H}}(\mathbb{F}_{q^m}^n, r) = \sum_{i=0}^{r} \binom{n}{i} (q^m - 1)^i \sim \begin{cases} \binom{n}{r} q^{rm} & \text{as } q \to +\infty \\ \binom{n}{r} q^{rm} & \text{as } m \to +\infty \\ \binom{n}{r} (q^m - 1)^r & \text{as } n \to +\infty \end{cases}$$

Theorem

- *Non-linear MDS codes are sparse:*

$$\lim_{q,n \to +\infty} \delta_q^H(\mathbb{F}_q^n, q^{n-d+1}, 0, d) = 0$$

- *Dense sub-linear MDS codes:*

$$\lim_{q,\ell \to +\infty} \delta_q^H(\mathbb{F}_{q^m}^n, q^{m(n-d+1)}, \ell, d) = 1$$

- *Sparse sub-linear MDS codes ($s = m/\ell$):*

$$\lim_{n,s \to +\infty} \delta_q^H(\mathbb{F}_{q^m}^n, q^{m(n-d+1)}, \ell, d) = 0$$

Rank Metric

$$D^{\mathrm{rk}}(u,v) := \dim_q \langle u_1 - v_1, \ldots, u_n - v_n \rangle_q$$

Singleton bound (MRD):

$$\frac{k\ell}{\max(m,n)} \leq \min(m,n) - d + 1$$

(for $\mathbb{F}_{q^\ell}$-linear codes in $\mathbb{F}_{q^m}^n$)

Volume of balls:

$$\mathbf{v}_q^{\mathrm{rk}}(\mathbb{F}_{q^m}^n, r) = \sum_{i=0}^{r} \begin{bmatrix} n \\ i \end{bmatrix}_q \prod_{j=0}^{i-1} (q^m - q^j)$$

Rank Metric Asymptotics

Theorem

- *Non-linear MRD codes are sparse:*

$$\lim_{q,n,m\to+\infty} \delta_q^{\mathrm{rk}}(\mathbb{F}_{q^m}^n, q^{\max\{n,m\}(\min\{n,m\}-d+1)}, 0, d) = 0.$$

- *Sub-linear (quasi-)MRD codes depend on the linearity degree:*

$\ell s \geq n$:

$$\lim_{q\to+\infty} \delta_q^{rk}(\mathbb{F}_{q^{\ell s}}^n, q^{\ell s(n-d+1)}, \ell, d) = \begin{cases} 1 & \text{if } \ell > (d-1)(n-d+1), \\ 0 & \text{if } \ell < (d-1)(n-d+1). \end{cases}$$

$\ell s < n$:

$$\lim_{q\to+\infty} \delta_q^{rk}(\mathbb{F}_{q^{\ell s}}^n, q^{\ell k}, \ell, d) = \begin{cases} 1 & \text{if } \ell > (d-1)(\ell s-d+1)+r, \\ 0 & \text{if } \ell < (d-1)(\ell s-d+1)+r. \end{cases}$$

where $k := \lfloor n(\ell s-d+1)/\ell \rfloor$ *and* $r := n(d-1) - \ell\lceil n(d-1)/\ell \rceil$.

Rank Metric Asymptotics II

Theorem

- *Density for growing linearity degree:*

$$\lim_{\ell \to +\infty} \delta_q^{rk}(\mathbb{F}_{q^{\ell s}}^n, q^{\ell s(n-d+1)}, \ell, d) = 1,$$

- *Bound for growing length:*

$$\limsup_{n \to +\infty} \delta_q^{rk}(\mathbb{F}_{q^{\ell s}}^n, q^{\ell \lfloor n(\ell s - d + 1)/\ell \rfloor}, \ell, d) \leq \frac{1}{1 + \begin{bmatrix} m \\ d-1 \end{bmatrix}_q q^{-2\ell}} < 1,$$

- *Bound for growing extension degree:*

$$\limsup_{s \to +\infty} \delta_q^{rk}(\mathbb{F}_{q^{\ell s}}^n, q^{\ell s(n-d+1)}, \ell, d) \leq \frac{q^\ell}{q^\ell + \begin{bmatrix} n \\ d-1 \end{bmatrix}_q} < 1.$$

Sum-Rank Metric

Ambient space ($n = t\eta$):

$$C \subseteq (\mathbb{F}_{q^m}^{\eta})^t$$

$$D^{\mathrm{srk}}(u, v) := \sum_{i=1}^{t} D^{\mathrm{srk}}(u^{(i)} - v^{(i)})$$

Singleton-type bound:

$$\frac{k\ell}{\max\{m, \eta\}} \leq t \min\{m, \eta\} - d + 1$$

Volume of balls:

$$\mathbf{v}_q^{\mathrm{sr,t}}(\mathbb{F}_{q^m}^n, r) = \sum_{h=0}^{r} \sum_{u \in U_h} \prod_{i=1}^{t} \begin{bmatrix} \eta \\ u_i \end{bmatrix}_q \prod_{j=0}^{u_i - 1} (q^m - q^j)$$

Sum-Rank Metric Asymptotics I

Theorem

Non-linear MSRD codes are sparse:

$$\lim_{q\to+\infty} \delta_q^{\mathrm{sr,t}}(\mathbb{F}_{q^m}^n, q^{\max\{m,\eta\}(t\min\{m,\eta\}-d+1)}, 0, d) = 0$$

Sum-Rank Metric Asymptotics I

Theorem

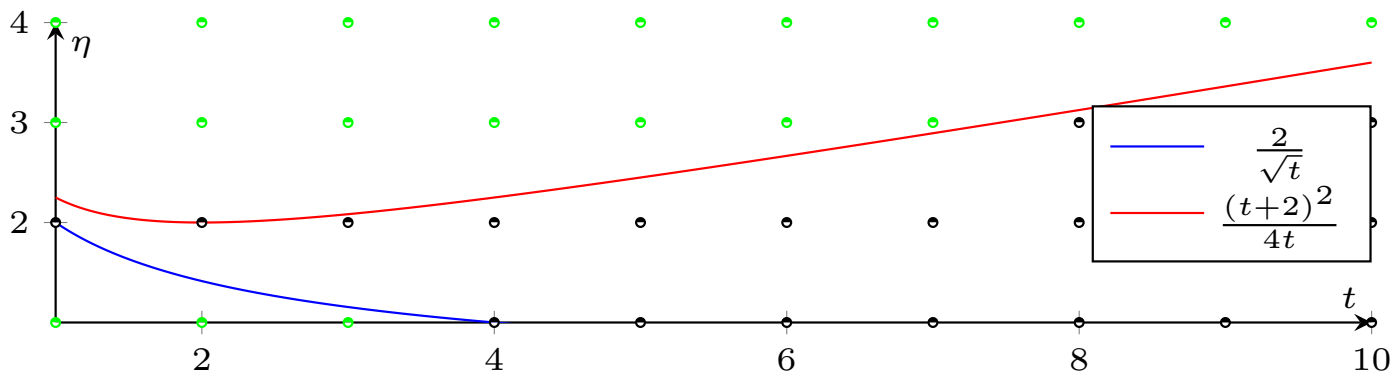
Non-linear MSRD codes are sparse:

$$\lim_{q\to+\infty} \delta_q^{\mathrm{sr,t}}(\mathbb{F}_{q^m}^n, q^{\max\{m,\eta\}(t\min\{m,\eta\}-d+1)}, 0, d) = 0$$

Corollary

Let $m \geq \eta$ and $2 \leq d \leq n$ be integers. We have

$$\lim_{q\to+\infty} \delta_q^{sr,t}(\mathbb{F}_{q^m}^n, q^{m(n-d+1)}, 1, d) = \begin{cases} 1 & \text{if } \eta < \frac{2}{\sqrt{t}} \\ 0 & \text{if } \eta > \frac{(t+2)^2}{4t}. \end{cases}$$



Sum-Rank Metric Asymptotics II

Theorem

Let $\theta := (d-1)\left(\min\{m,\eta\} - \frac{d-1}{t}\right) + \frac{\tilde{z}^2}{t} - \tilde{z}$.

- *If* $m \geq \eta$, *then:*

$$\lim_{q\to+\infty} \delta_q^{sr,t}(\mathbb{F}_{q^m}^n, q^{m(n-d+1)}, \ell, d) = \begin{cases} 1 & \text{if } \theta < \ell, \\ 0 & \text{if } \theta > \ell \end{cases}$$

- *If* $\eta > m$, *then:*

$$\lim_{q\to+\infty} \delta_q^{sr,t}(\mathbb{F}_{q^m}^n, q^{\ell k}, \ell, d) = \begin{cases} 1 & \text{if } \theta - r < \ell, \\ 0 & \text{if } \theta - r > \ell, \end{cases}$$

where $k = \left\lfloor \frac{\eta(mt-d+1)}{\ell} \right\rfloor$ *and* $r = \ell\left(\left\lceil \frac{\eta(d-1)}{\ell} \right\rceil - \frac{\eta(d-1)}{\ell}\right)$.

So...

- Random non-linear codes have an expected minimum distance (way) below the GV-bound.
- Random linear codes have an expected minimum distance of at least the GV-bound.
- If the linearity degree grows with the field size, then random codes are optimal by high probability.

basically independent of the metric

1 The Naive Expectation

2 Known Bounds

3 Behavior of Random Codes (Genericity)

4 Linearity as a Constraint

5 The Big Picture

Known constructions of non-linear codes

Some non-linear binary codes have better parameters than linear:

- Kerdock codes
- Preparata codes
- Goethals codes
- Delsarte-Goethals codes
- Nordstrom-Robinson code

Known constructions of non-linear codes

Some non-linear binary codes have better parameters than linear:

- Kerdock codes
- Preparata codes
- Goethals codes
- Delsarte-Goethals codes
- Nordstrom-Robinson code

All of these arise as Gray images of linear $\mathbb{Z}_4$-codes!

Known constructions of non-linear codes

Some non-linear binary codes have better parameters than linear:

- Kerdock codes
- Preparata codes
- Goethals codes
- Delsarte-Goethals codes
- Nordstrom-Robinson code

All of these arise as Gray images of linear $\mathbb{Z}_4$-codes!

Also many other sporadic examples turn out to be $\mathbb{Z}_{2^s}$-linear or sub-linear/additive.

MDS codes

For a code of length n, size M, and distance d,

$$M \leq q^{n-d+1}$$

- Singleton bound: does *not* assume linearity.
- Reed–Solomon codes attain the bound.
- MDS conjecture (now a theorem for many parameters): linear MDS codes require roughly $q \geq n - 1$.

MDS codes

For a code of length n, size M, and distance d,

$$M \leq q^{n-d+1}$$

- Singleton bound: does *not* assume linearity.
- Reed–Solomon codes attain the bound.
- MDS conjecture (now a theorem for many parameters): linear MDS codes require roughly $q \geq n - 1$.

- No non-linear code is known that violates the same length restrictions.
- All known non-linear MDS constructions are equivalent to linear MDS codes.

MDS codes

For a code of length n, size M, and distance d,

$$M \leq q^{n-d+1}$$

- Singleton bound: does *not* assume linearity.
- Reed–Solomon codes attain the bound.
- MDS conjecture (now a theorem for many parameters): linear MDS codes require roughly $q \geq n - 1$.

- No non-linear code is known that violates the same length restrictions.
- All known non-linear MDS constructions are equivalent to linear MDS codes.

For MDS codes, non-linearity appears to buy nothing.

A^sMDS codes and the maximal arc bound

Singleton defect:

$$s = n - k + 1 - d.$$

- $s = 0$: MDS codes.
- $s = 1$: AMDS codes.
- More generally: A^sMDS codes.

A^sMDS codes and the maximal arc bound

Singleton defect:

$$s = n - k + 1 - d.$$

- $s = 0$: MDS codes.
- $s = 1$: AMDS codes.
- More generally: A^sMDS codes.

Maximal arcs bound (for linear codes):

$$n \leq (s+1)(q+1) + k - 2.$$

Codes attaining equality are called *length-maximal*.

A^sMDS codes and the maximal arc bound

Singleton defect:

$$s = n - k + 1 - d.$$

- $s = 0$: MDS codes.
- $s = 1$: AMDS codes.
- More generally: A^sMDS codes.

Maximal arcs bound (for linear codes):

$$n \leq (s+1)(q+1) + k - 2.$$

Codes attaining equality are called *length-maximal*.

Recent results: For most parameter regimes the same bound holds without assuming linearity. (For a few parameters a gap still exists.)

Weight distribution constraints

For linear codes, the MacWilliams identity

- relates weight enumerators of C and $C^{\perp}$,
- imposes strong global constraints on distance distributions.

It can be extended to sub-linear or $\mathbb{Z}_{2^2}$-linear codes.

Weight distribution constraints

For linear codes, the MacWilliams identity

- relates weight enumerators of C and $C^{\perp}$,
- imposes strong global constraints on distance distributions.

It can be extended to sub-linear or $\mathbb{Z}_{2^2}$-linear codes.

On the other hand, even for non-linear codes we have restrictions on the distance distribution by

- association schemes,
- LP bounds,
- geometry/design theory.

Code constructions

- For small parameters, better-than-linear non-linear codes for small parameters are known. ($\rightarrow$ Brouwer's tables)
- General families of non-linear codes have turned out to be equivalent to linear codes.
- Mathon's non-linear $(21, 9^3, 18)_9$ AMDS code is not equivalent to a linear code, but it is sub-linear.

1 The Naive Expectation

2 Known Bounds

3 Behavior of Random Codes (Genericity)

4 Linearity as a Constraint

5 The Big Picture

The big picture

- Linearity is a constraint $\implies$ theoretically more possibilities for good non-linear codes.
- Sporadic examples of better-than-linear non-linear codes exist.

The big picture

- Linearity is a constraint $\implies$ theoretically more possibilities for good non-linear codes.
- Sporadic examples of better-than-linear non-linear codes exist.
- All known general families of good non-linear codes are sub-linear or equivalent to linear.

The big picture

- Linearity is a constraint $\implies$ theoretically more possibilities for good non-linear codes.
- Sporadic examples of better-than-linear non-linear codes exist.
- All known general families of good non-linear codes are sub-linear or equivalent to linear.
- Good non-linear codes are sparse.

The big picture

- Linearity is a constraint $\implies$ theoretically more possibilities for good non-linear codes.
- Sporadic examples of better-than-linear non-linear codes exist.
- All known general families of good non-linear codes are sub-linear or equivalent to linear.
- Good non-linear codes are sparse.
- $\mathbb{F}_{q^\ell}$-linear codes attain GV-bound for growing q or $\ell = m/s$.
- The same for fixed-rate codes arbitrarily close to the GV-bound for growing n or s.

The big picture

- Linearity is a constraint $\implies$ theoretically more possibilities for good non-linear codes.
- Sporadic examples of better-than-linear non-linear codes exist.
- All known general families of good non-linear codes are sub-linear or equivalent to linear.
- Good non-linear codes are sparse.
- $\mathbb{F}_{q^\ell}$-linear codes attain GV-bound for growing q or $\ell = m/s$.
- The same for fixed-rate codes arbitrarily close to the GV-bound for growing n or s.
- For Singleton-type bound it depends on the metric and the linearity degree. High linearity degree leads to optimality.

The big picture

Good nonlinear are sporadic.
Good linear codes are generic.

- Linearity is a constraint $\implies$ theoretically more possibilities for good non-linear codes.
- Sporadic examples of better-than-linear non-linear codes exist.
- All known general families of good non-linear codes are sub-linear or equivalent to linear.
- Good non-linear codes are sparse.
- $\mathbb{F}_{q^\ell}$-linear codes attain GV-bound for growing q or $\ell = m/s$.
- The same for fixed-rate codes arbitrarily close to the GV-bound for growing n or s.
- For Singleton-type bound it depends on the metric and the linearity degree. High linearity degree leads to optimality.

Thank you for your attention!
Questions? – Comments?

