Semifields, and their relations to cryptography

Lukas Kölsch

University of South Florida

02/13/2023

Public Key Cryptography today

With the quantum computer on the horizon, new public key cryptosystems have to be designed.

Many new ideas are developed, based on e.g. *lattices, codes, isogenies, systems of equations.*

Public Key Cryptography today

With the quantum computer on the horizon, new public key cryptosystems have to be designed.

Many new ideas are developed, based on e.g. *lattices*, *codes*, *isogenies*, *systems of equations.*

Codes

Classical coding theory is largely concerned with linear codes over a finite field.

Definition (Code)

A linear $[n, k, d]$-code $\mathcal{C} \leq \mathbb{F}_q^n$ is a k-dimensional subspace of $\mathbb{F}_q^n$ such that

$$d = \min_{x,y \in \mathcal{C}, x \neq y} \#\{j \in \{1, \dots, n\} \colon x_j \neq y_j\}.$$

The distance used is the *Hamming distance*. Classical problems:

- Find codes that are "optimal" (achieve maximal d with fixed n, k)
- Find for a given point $x \in \mathbb{F}_q^n$ the point in $\mathcal{C}$ closest to x. (decoding problem)

Cryptography via Coding Theory

Definition (McEliece cryptosystem - High level view)

- Alice (recipient) chooses a code with good parameters and efficient decoding
- Alice "scrambles" the code and publishes the scrambled code. Attackers are not able to "unscramble" the code.
- Bob (sender) encodes his message with the public, scrambled code and adds an error.
- Alice can "unscramble" the code and use the efficient decoding algorithm.
- An attacker has no access to an efficient decoding algorithm because they cannot unscramble the code.

Cryptography via Coding Theory

In order for the McEliece cryptosystem to work:

- We must know families of good codes with efficient decoding algorithms
- The attacker should not be able to "unscramble" the code
- decoding of the "scrambled" code should be *difficult*.

To ensure this, the code has to be large enough, resulting in large key sizes.
This is the *major drawback* of code-based cryptography!

A new approach

If decoding of a random code was harder, we could use smaller codes and thus smaller keys. We can use *non-classical* coding theory like rank-metric codes!

Definition (Rank-metric code)

Let $M_{n,m}(\mathbb{F}_q)$ the set of $n \times m$-matrices over $\mathbb{F}_q$. A linear *rank-metric code* is a subspace $\mathcal{C} \leq M_{n,m}(\mathbb{F}_q)$ with minimum distance

$$d = \min_{X,Y \in \mathcal{C}, X \neq Y} \operatorname{rk}(X - Y).$$

The distance used here is the *rank metric*.

Decoding in the rank metric (seems to be) harder than in the Hamming metric.

McEliece in the rank metric

We can adapt the McEliece cryptosystem to rank-metric codes without any difficulties.

But: We need to find good *rank-metric* codes!

2 of 7 code-based cryptosystems in the 2nd round of the recent NIST post-quantum cryptography competition were McEliece-like using rank-metric codes.

NIST encouraged further work on rank-metric code based cryptography.

Constructions of rank-metric codes

What is a *good* rank-metric code?

Theorem (Singleton-like bound, Delsarte 1978)

Suppose $\mathcal{C} \leq M_{n,m}(\mathbb{F}_q)$ with minimum distance d. Then

$$|\mathcal{C}| \leq q^{n(m-d+1)}.$$

Rank-metric codes satisfying the bound with equality are called maximum rank distance (MRD) code.

There are few constructions of MRD codes and even less have efficient decoding algorithms.

Most constructions of MRD codes are related to an algebraic structure called *semifield*.

Semifields

Definition

A (finite) semifield $\mathbb{S} = (S, +, \circ)$ is a finite set S equipped with two operations $(+, \circ)$ satisfying the following axioms.

(S1) $(S, +)$ is a group.

(S2) For all $x, y, z \in S$,

- $x \circ (y + z) = x \circ y + x \circ z$,
- $(x + y) \circ z = x \circ z + y \circ z$.

(S3) For all $x, y \in S$, $x \circ y = 0$ implies $x = 0$ or $y = 0$.

Basic properties

If $\circ$ is associative then $\mathbb{S}$ is essentially a finite field (Wedderburn's Theorem).

The additive group of a semifield $(\mathbb{S}, +, \circ)$ is always an elementary abelian p-group.

We can thus identify the additive group of a semifield $\mathbb{S}$ with the additive group of the finite field $\mathbb{F}_{p^n}$.

Semifields were studied in pure mathematics for decades because of connections to finite geometry and difference sets.

Connecting semifields and MRD codes

We have a simple key connection.

Theorem

Let $\mathbb{S} = (\mathbb{F}_q^n, +, \circ)$ *be a semifield. Then the set of left-multiplications*

$$L_x(y) = x \circ y,\ \mathcal{C} = \{L_x \colon x \in \mathbb{F}_q^n\}$$

defines a linear MRD code with parameters $d = m = n$.

Note: The distributivity law $x \circ (y + z) = x \circ y + x \circ z$ implies that L_x is a linear mapping.

The MRD codes constructed by semifields are *square, full rank* MRD codes.

Connecting semifields and MRD codes

Even more:

Theorem (de la Cruz, Kiermaier, Wassermann, Willems, 2015)

There is a 1-1 correspondence between finite semifields and linear, square full rank MRD codes.

Connecting semifields and MRD codes

Even more:

Theorem (de la Cruz, Kiermaier, Wassermann, Willems, 2015)

There is a 1-1 correspondence between finite semifields and linear, square full rank MRD codes.

And even more: Almost all other known constructions of MRD codes start from a square full rank MRD code - and are thus connected to semifields.

Connecting semifields and MRD codes

The oldest known MRD code is the *Gabidulin code* .

Gabidulin codes $\approx$ Reed-Solomon codes in the rank metric.

They are related to the full rank MRD code obtained by choosing $\mathbb{S} = \mathbb{F}_{q^n}$.

There is an efficient decoding algorithm for Gabidulin codes!

McEliece style cryptosystems based on Gabidulin codes have been proposed (e.g. Gabidulin 1991, Loidreau 2017), but many were broken.

Connecting semifields and MRD codes

Problem: Gabidulin codes have a lot of structure (like invariant subspaces).

Ideas:

- Tweak the McEliece system with Gabidulin codes (e.g. RQC proposal to NIST)
- Find other MRD (or almost-MRD) codes that can be decoded efficiently

Connecting semifields and MRD codes

Problem: Gabidulin codes have a lot of structure (like invariant subspaces).

Ideas:

- Tweak the McEliece system with Gabidulin codes (e.g. RQC proposal to NIST)
- Find other MRD (or almost-MRD) codes that can be decoded efficiently

Since Gabidulin codes are related to $\mathbb{S} = \mathbb{F}_{q^n}$, other semifields might give similar codes!

Connecting semifields and MRD codes

Goal

Construct more semifields and investigate the resulting MRD codes.

Things to look out for:

- *Commutative* semifields yield MRD codes that can be stored as symmetric matrices
- Different semifields might yield equivalent codes

Definition (Isotopy)

Two semifields $\mathbb{S}_1 = (\mathbb{F}_p^n, +, \circ_1)$ and $\mathbb{S}_2 = (\mathbb{F}_p^n, +, \circ_2)$ are *isotopic* if there exist $\mathbb{F}_p$-linear bijections L, M and N of $\mathbb{F}_{p^n}$ satisfying

$$N(x \circ_1 y) = L(x) \circ_2 M(y).$$

Such a triple $\gamma = (N, L, M)$ is called an *isotopism* between $\mathbb{S}_1$ and $\mathbb{S}_2$.

Definition (Autotopism and the Autotopism group)

The autotopism group $\mathrm{Aut}(\mathbb{S})$ of a semifield $\mathbb{S} = (\mathbb{F}_p^n, +, \circ)$ is defined by

$$\mathrm{Aut}(\mathbb{S}) = \{(N, L, M) \in \mathrm{GL}(\mathbb{F}_{p^n})^3 \colon N(x \circ y) = L(x) \circ M(y)\}.$$

Two semifields are isotopic iff the associated rank-metric codes are equivalent.

Examples of families of semifields

Example (Twisted fields, Albert, 1961)

Let $K = \mathbb{F}_{p^n}$, $n > 2$, and define $\circ\colon K \to K$ via

$$x \circ y = xy - ax^q y^r,$$

where $a \notin \mathbb{F}_{p^n}^{q-1} \cdot \mathbb{F}_{p^n}^{r-1}$ and q, r are powers of p. Then $\mathbb{S} = (K, +, \circ)$ is a semifield.

The twisted fields yield *twisted Gabidulin codes* as MRD codes (Sheekey, 2015).

Efficient decoding algorithms for twisted Gabidulin codes have been found (e.g. Randrianarisoa, Rosenthal, 2017) and have been proposed for use in McEliece-like cryptosystems.

Examples of families of semifields

Example (Dickson, 1905)

Let $K = \mathbb{F}_{p^m} \times \mathbb{F}_{p^m}$ with p odd and define $\circ \colon K \times K \to K$ via

$$(x, y) \circ (u, v) = (xu + a(yv)^q, xv + yu),$$

where a is a non-square in $\mathbb{F}_{p^m}$ and q is a power of p. Then $\mathbb{S} = (K, +, \circ)$ is a (commutative) semifield.

This is a *bivariate construction*.

There are many bivariate constructions!

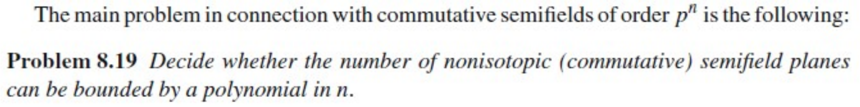
The known commutative semifields of size p^n, p odd

Until 2022:

Family	**Count**	Proven in	Bivariate?
The finite field	1	trivial	$\approx$ Yes
Dickson	$\approx n/4$	1905	Yes
Albert's twisted Fields	$\approx n/2$	1961	$\approx$ Yes
Ganley	1 ($p = 3$ only)	1981	No
Cohen-Ganley	1 ($p = 3$ only)	1982	No
Coulter-Matthews-Ding-Yuan	2 ($p = 3$ only)	2006	No
Zha-Kyureghyan-Wang	??	2008	No
Budaghyan-Helleseth	$\approx n/2$	2009	Yes
Bierbrauer$_3$	$\approx n/2$	2010	No
Bierbrauer$_4$	$\approx n/2$	2010	No
Zhou-Pott	$\approx n^2$	2013	Yes

The known commutative semifields of size p^n, p odd

Open problem!



The main problem in connection with commutative semifields of order p^n is the following:

Problem 8.19 *Decide whether the number of nonisotopic (commutative) semifield planes can be bounded by a polynomial in n.*

Pott, A.: Almost perfect and planar functions, *Designs, Codes, Cryptography* (2016)

Bivariate constructions

We are interested in special bivariate constructions where $K = \mathbb{F}_{p^m} \times \mathbb{F}_{p^m}$ and

$$(x, y) \circ (u, v) = (f(x, y, u, v), g(x, y, u, v)),$$

and f, g are homogeneous of degree $q + 1$ (resp. $r + 1$) where q, r are powers of p.

Example (Göloglu, K., 2022)

Let $K = \mathbb{F}_{p^m} \times \mathbb{F}_{p^m}$, m even and set

$$(x, y) \circ (u, v) = (x^q u + xu^q + b(y^q v + yv^q), x^r v + yu^r + a/b(yv^r + y^r v)),$$

where p odd, $q = p^k$, $r = p^{k+m/2}$, $b \in \mathbb{F}_{p^m}$ is a non-square, $a \in \mathbb{F}^*_{p^{m/2}}$, $m/\gcd(k, m)$ is odd.

Why this structure?

We are interested in special bivariate constructions where $K = \mathbb{F}_{p^m} \times \mathbb{F}_{p^m}$ and

$$(x, y) \circ (u, v) = (f(x, y, u, v), g(x, y, u, v)),$$

and f, g are homogeneous of degree $q + 1$ (resp. $r + 1$) where q, r are powers of p.

Why this structure?

We are interested in special bivariate constructions where $K = \mathbb{F}_{p^m} \times \mathbb{F}_{p^m}$ and

$$(x, y) \circ (u, v) = (f(x, y, u, v), g(x, y, u, v)),$$

and f, g are homogeneous of degree $q + 1$ (resp. $r + 1$) where q, r are powers of p.

These semifields have some nice autotopisms! Namely, if $L = M = \left(\begin{smallmatrix} a & 0 \\ 0 & a \end{smallmatrix}\right)$ then

$$L(x, y) \circ M(u, v) = (a^{q+1} f(x, y, u, v), a^{r+1} g(x, y, u, v)),$$

so (N, L, M) with $N = \begin{pmatrix} a^{q+1} & 0 \\ 0 & a^{r+1} \end{pmatrix}$ is an autotopism for any $a \in \mathbb{F}_{p^m}^{\times}$.

$\implies$ These semifields always have a cyclic subgroup in their autotopism group of order $p^m - 1$.

Why this structure?

Another reason is: It turns out many of the known bivariate semifields semifields "secretly" have this structure!

Example (Zhou-Pott, 2013)

Let $K = \mathbb{F}_{p^m} \times \mathbb{F}_{p^m}$.

$$(x, y) \circ (u, v) = (x^q u + u^q x + \alpha(y^q v + yv^q)^r, xv + yu)$$

where $q = p^k$, $r = p^l$, $\gcd(k, m)/m$ is odd, and α is a non-square in $\mathbb{F}_{p^m}$.

..is isotopic to...

$$(x, y) \circ (u, v) = (x^q u + u^q x + \alpha(y^q v + yv^q), x^r v + yu^r).$$

And many more (e.g. Dickson, Budaghyan-Helleseth....)!

What can we do with this structure?

1. Systematically search for new semifields that have this structure.

2. Use the nice subgroup in the autotopism group to answer the isotopy question!

Isotopy of semifields

Question

How can we decide if different semifields are isotopic or not? Can we count the (known) semifields up to isotopy?

Example (Göloglu, K., 2022)

Let $K = \mathbb{F}_{p^m} \times \mathbb{F}_{p^m}$, m even and set

$$(x, y) \circ (u, v) = (x^q u + xu^q + b(y^q v + yv^q), x^r v + yu^r + a/b(yv^r + y^r v)),$$

where p odd, $q = p^k$, $r = p^{k+m/2}$, $b \in \mathbb{F}_{p^m}$ is a non-square, $a \in \mathbb{F}^*_{p^{m/2}}$, $m/\gcd(k, m)$ is odd.

Which choices for q, a, b yield non-isotopic semifields? This is in general a very hard question!

Isotopy via the autotopism group

Lemma

Assume $\mathbb{S}_1$,$\mathbb{S}_2$ *are isotopic semifields of order* p^n. *Then* $\mathrm{Aut}(\mathbb{S}_1)$ *and* $\mathrm{Aut}(\mathbb{S}_2)$ *are conjugate in* $\mathrm{GL}(\mathbb{F}_{p^n})^3$.

Problem: Determining the autotopism group is also very hard!

There is sometimes a way to use the lemma without knowing the autotopism group - if one can identify a large and nice subgroup first.

Recall our bivariate semifields have a cyclic subgroup of order $p^m - 1$ in the autotopism group!

Show that two bivariate semifields $\mathbb{S}_1, \mathbb{S}_2$ are not isotopic - in five simple steps!

- Let $H_1 \leq \operatorname{Aut}(\mathbb{S}_1)$, $H_2 \leq \operatorname{Aut}(\mathbb{S}_2)$ with $|H_1| = |H_2| = p^m - 1$ be the nice cyclic autotopism subgroups.
- Choose a suitable prime p' and Sylow p'-groups $S_1 \leq H_1$, $S_2 \leq H_2$.
- Prove that S_1, S_2 are also Sylow p'-groups of $\operatorname{Aut}(\mathbb{S}_1), \operatorname{Aut}(\mathbb{S}_2)$ (key step!)
- If $\gamma^{-1} \operatorname{Aut}(\mathbb{S}_1)\gamma = \operatorname{Aut}(\mathbb{S}_2)$ then $\gamma^{-1} S_1 \gamma$ is a Sylow subgroup of $\operatorname{Aut}(\mathbb{S}_2)$. So $\gamma^{-1} S_1 \gamma$ and S_2 are conjugate in $\operatorname{Aut}(\mathbb{S}_2)$ (by Sylow's theorem)!
- Determine all $\delta \in \operatorname{GL}(\mathbb{F}_{p^n})^3$ such that $\delta^{-1} S_1 \delta = S_2$. If all $\delta \notin \operatorname{Aut}(\mathbb{S}_2)$ then S_1, S_2 are not isotopic.

In some sense, checking $\gamma^{-1} \operatorname{Aut}(\mathbb{S}_1)\gamma = \operatorname{Aut}(\mathbb{S}_2)$ is reduced to checking $\delta^{-1} S_1 \delta = S_2$.

From this procedure we get the following result:

Theorem (Göloğlu, K., 2022)

If two sufficiently nice bivariate semifields defined over $\mathbb{F}_{p^m} \times \mathbb{F}_{p^m}$ are isotopic then there exists an isotopism $\gamma = (N, L, M) \in \Gamma L(2, \mathbb{F}_{p^m})^3$ between them.

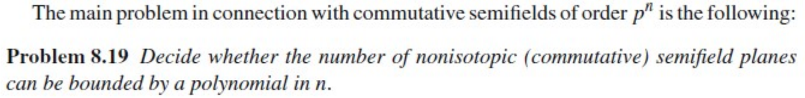
This simplifies the isotopy question for all nice bivariate semifields.

Isotopisms of the form $\gamma = (N, L, M) \in \Gamma L(2, \mathbb{F}_{p^m})^3$ are (comparatively) easy to determine.

The known commutative semifields of size p^n, p odd

Family	**Count**	Proven in	Bivariate?
The finite field	1	trivial	$\approx$ Yes
Dickson	$\approx n/4$	1905	Yes
Albert's twisted Fields	$\approx n/2$	1961	$\approx$ Yes
Ganley	1 ($p = 3$ only)	1981	No
Cohen-Ganley	1 ($p = 3$ only)	1982	No
Coulter-Matthews-Ding-Yuan	2 ($p = 3$ only)	2006	No
Zha-Kyureghyan-Wang	??	2008	No
Budaghyan-Helleseth	$\approx n/2$	2009	Yes
Bierbrauer$_3$	$\approx n/2$	2010	No
Bierbrauer$_4$	$\approx n/2$	2010	No
Zhou-Pott	$\approx n^2$	2013	Yes
Göloğlu-K.	$\approx p^{n/4}$	2022	Yes

The known commutative semifields of size p^n, p odd

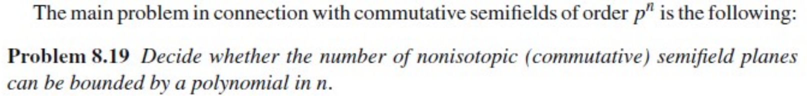


The main problem in connection with commutative semifields of order p^n is the following:

Problem 8.19 *Decide whether the number of nonisotopic (commutative) semifield planes can be bounded by a polynomial in n.*

Pott, A.: Almost perfect and planar functions, *Designs, Codes, Cryptography* (2016)

The known commutative semifields of size p^n, p odd



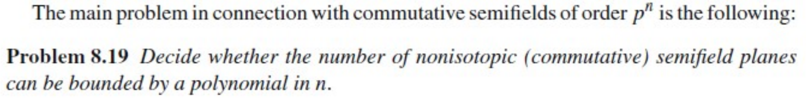
The main problem in connection with commutative semifields of order p^n is the following:

Problem 8.19 *Decide whether the number of nonisotopic (commutative) semifield planes can be bounded by a polynomial in n.*

Pott, A.: Almost perfect and planar functions, *Designs, Codes, Cryptography* (2016)

This problem is now solved!

The known commutative semifields of size p^n, p odd



The main problem in connection with commutative semifields of order p^n is the following:

Problem 8.19 *Decide whether the number of nonisotopic (commutative) semifield planes can be bounded by a polynomial in n.*

Pott, A.: Almost perfect and planar functions, *Designs, Codes, Cryptography* (2016)

This problem is now solved!

This also yields the biggest family of commutative MRD codes!

The known non-commutative semifields of size p^n, p odd

Non-commutative semifields:

- There are more constructions, e.g. via skew-polynomial rings (Petit, 1966), finite geometry (Jha, Johnson, 1990) or secondary constructions based on commutative semifields
- However, counting (up to isotopy) is much more difficult!
- Several families have $\approx p^{n/2}$ non-isotopic elements (Kantor 2003, Lavrauw 2013, Sheekey 2019)

The known non-commutative semifields of size p^n, p odd

Non-commutative semifields:

- There are more constructions, e.g. via skew-polynomial rings (Petit, 1966), finite geometry (Jha, Johnson, 1990) or secondary constructions based on commutative semifields
- However, counting (up to isotopy) is much more difficult!
- Several families have $\approx p^{n/2}$ non-isotopic elements (Kantor 2003, Lavrauw 2013, Sheekey 2019)
- The "square-root barrier" was broken in (Göloğlu, K. , 2023+). We presented a family with $\approx p^{2n/3}$ non-isotopic semifields.

Semifields and symmetric cryptography

To resist differential attacks, a block cipher needs to be nonlinear.

Definition

A function $F\colon \mathbb{F}_{p^n} \to \mathbb{F}_{p^n}$ is called *perfect nonlinear* if the equation (in x)

$$F(x+a) - F(x) = b$$

has exactly one solution for any b and any non-zero a.

Theorem (Coulter, Henderson, 2007)

If $\mathbb{S} = (\mathbb{F}_p^n, +, \circ)$ is a commutative semifield, then $F(x) = x \circ x$ is perfect nonlinear.

Constructions of new commutative semifields give new perfect nonlinear functions.

Current and Future work

Rank-metric codes from semifields:

- Constructions of MRD codes based on new semifields we found.
- Adapting decoding algorithms of (twisted) Gabidulin codes to other bivariate semifields.
- Check if the new codes are resistant to attacks that broke Gabidulin based McEliece (e.g. Overbeck's attack)

Current and Future work

. . . on the more theoretical side.

Problem (Kantor's conjecture, 2003)

Prove that the number of non-isotopic semifields of odd order $N = p^n$ is at least exponential in N.

The best current bound is $p^{2n/3}$, not even linear in N.

Interestingly, in characteristic 2 a family with exponentially many semifields has been found (Kantor and Williams, 2005).

Thank you for your attention!

The talk is based on:

Göloğlu, F., Kölsch, L.: An exponential bound on the number of non-isotopic commutative semifields. *Transactions of the American Mathematical Society*, 2022.

Göloğlu, F., Kölsch, L.: Counting the number of non-isotopic Taniguchi semifields. To appear in *Designs, Codes, Cryptography*, 2023.

Göloğlu, F., Kölsch, L.: Equivalences of biprojective almost perfect nonlinear functions. Preprint, 2022.

. . . and ongoing projects with Faruk Göloğlu (Charles Univ. Prague), Jean-Francois Biasse, Giacomo Micheli (Univ. of South Florida).