Nonlinear Shift Registers: A Survey and Open Problems

Tor Helleseth

University of Bergen

NORWAY

Outline

- Introduction
- Nonlinear Shift Registers (NLFSRs)
  - Some basic theory
- De Bruijn Graph
  - De Bruijn graph
  - Golomb's conjecture/Mykkeltveit's proof
- Period of NLFRs
- Connections to Finite Fields
  - Cross-join pairs
  - Cycle-joining and cyclotomy

Linear Recursion

- Linear recurrence

$$s_{t+n} + c_{n-1} s_{t+n-1} + \ldots + c_0 s_t = 0, \quad c_i , s_i \in GF(p)$$

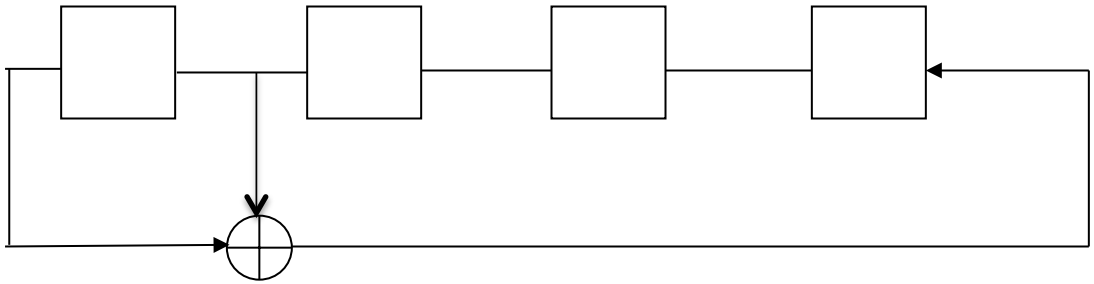
- Characteristic polynomial

$$f(x) = x^n + c_{n-1} x^{n-1} + \ldots + c_0$$

$\Omega(f)$ = The 2^n binary sequences generated by recursion

- Properties
  - "Easy" to find period of the sequences in $\Omega(f)$ from $f(x)$
    - Period determined by smallest e such that $f(x) \mid x^e - 1$
    - All sequences in $\Omega(f)$ have period e
    - Smallest period for at least one sequences in $\Omega(f)$
  - Bounds on the distribution of elements in (s_t) are evaluated using methods from finite fields

m-sequences



Linear recurrence

$$s_{t+4} + s_{t+1} + s_t = 0$$

Primitive polynomial

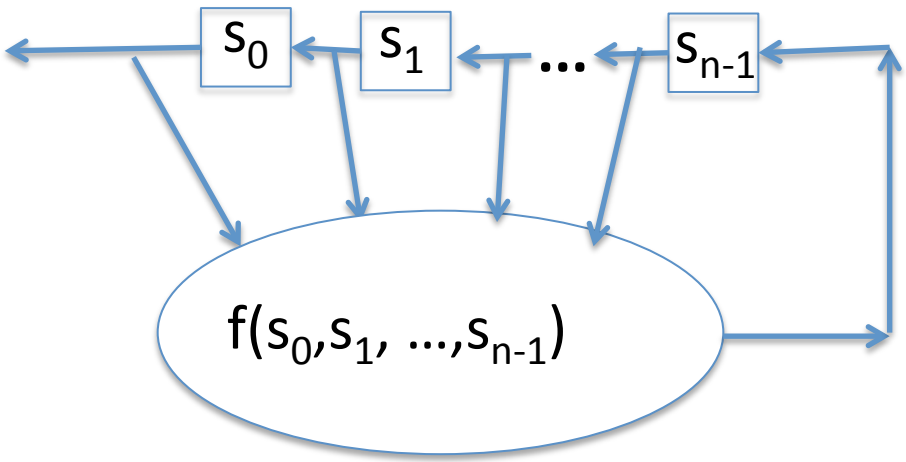
$$f(x) = x^4 + x + 1$$

(s_t) : 000100110101111 . . .

General properties of m-sequences

- Period $\varepsilon = 2^n - 1$
- Balanced (except for a missing 0)
- Run property
- $s_t - s_{t+\tau} = s_{t+\gamma}$, $s_{2t} = s_{t+\delta}$
- During a period all nonzero n-tuples occur

Nonlinear Shift Registers



- The feedback polynomial is nonlinear of the form

$$f(s_0,s_1,\ldots s_{n-1}) = \Sigma_{I\varepsilon\{0,1\}^n} c_I\, s_0^{i_0}\, s_1^{i_1} \ldots s_{n-1}^{i_{n-1}}$$

- Determined by truth table giving $f(s_0,s_1,\ldots s_{n-1})$ for all possible 2^n values
- Number of nonlinear polynomials (Boolean functions) in n variables is 2^{2^n}

Nonlinear Shift Registers - Challenges

Motivation

- NLSRs are used as building blocks in many modern stream ciphers (Grain, Trivium, Mickey, Pomaranch, …)
- Increase complexity of the key stream in stream ciphers

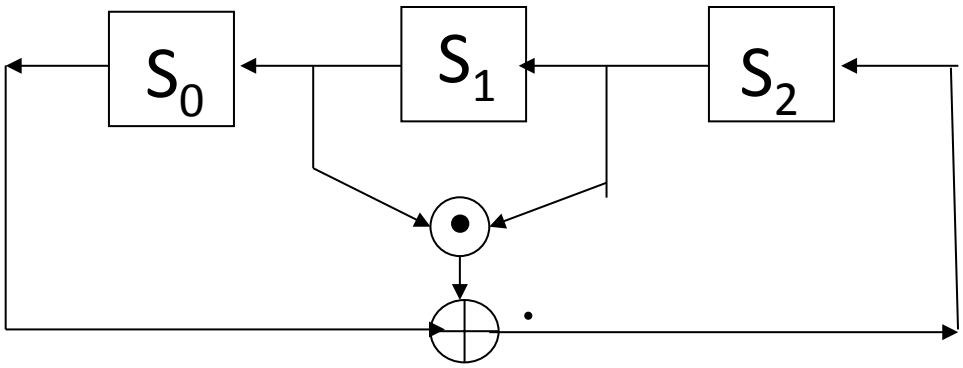
Challenges for NLFSRs

- How to determine the period of sequences from NLFSRs
- No general theory exists and many ad-hoc techniques have to be invented for these problems
- Constructing efficiently large classes long sequences of period 2^n (de Bruijn sequences)/Classify de Bruijn sequences
- Find algebraic methods to analyze NLFSRs
- Find the distribution of the elements in sequences generated by an NLFSR

Nonlinear Shift Register - Example

- A nonlinear recursion in n-variables can be described using its truth table (Example n=3)

s_0	s_1	s_2	$f(s_0\ s_1\ s_2)$
0	0	0	**0**
0	0	1	**0**
0	1	0	**0**
0	1	1	**1**
1	0	0	**1**
1	0	1	**1**
1	1	0	**1**
1	1	1	**0**



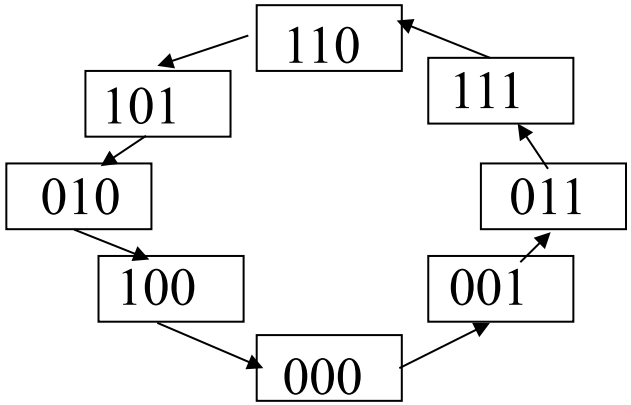
$$f(s_0,s_1,s_2) = s_0+s_1s_2$$

$$(s_{t+2} = s_t + s_{t+1}s_{t+2})$$

- The number of Boolean functions in n-variables are 2^{2^n}
- The number of linear Boolean functions are 2^n

Example – de Bruijn Sequence

- Let $f(s_0,s_1,s_2) = 1+s_0+s_1+s_1s_2$



- This gives a maximal sequence of length 2^n

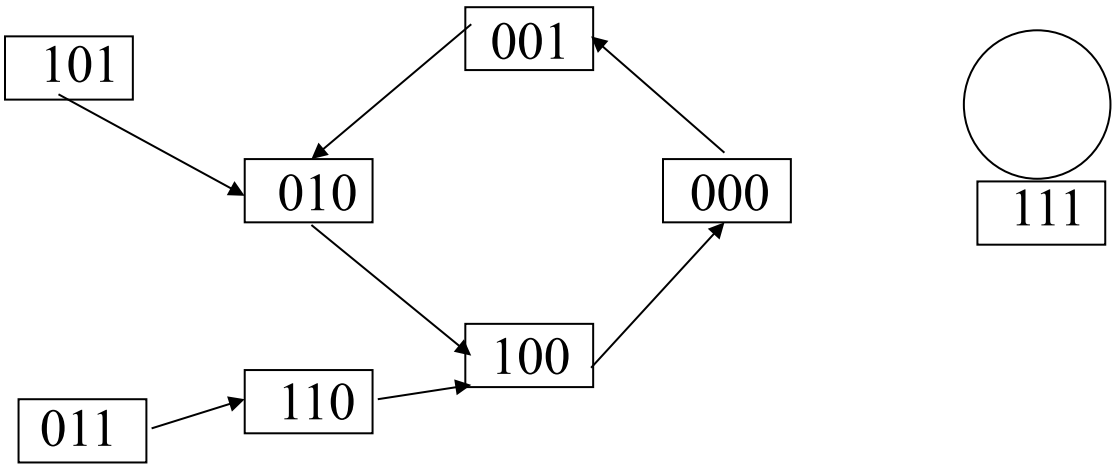
  … 11010001 …

  and is called a de Bruijn sequence

- Number of de Bruijn sequences of period 2^n are $2^{2^{n-1}-n}$

Example – Singular f

- Let $f(s_0,s_1,s_2) = 1+s_0+s_1+s_2+s_0s_1+s_0s_2+s_1s_2$



- Contains "branch point" and such an f is called singular
- f is nonsingular if and only if $f = s_0 + g(s_1,\ldots,s_{n-1})$
- Then $(s_0, s_1, \ldots, s_{n-1}) \rightarrow (s_1, s_2, \ldots, s_{n-1}, f(s_1,s_2,\ldots,s_{n-1}))$ is a permutation of B_n

De Bruijn Graph

- Directed graph
- 2^n nodes (states) $\leftrightarrow$ $(s_0, s_1, \ldots, s_{n-1})$
- Each state has two successors

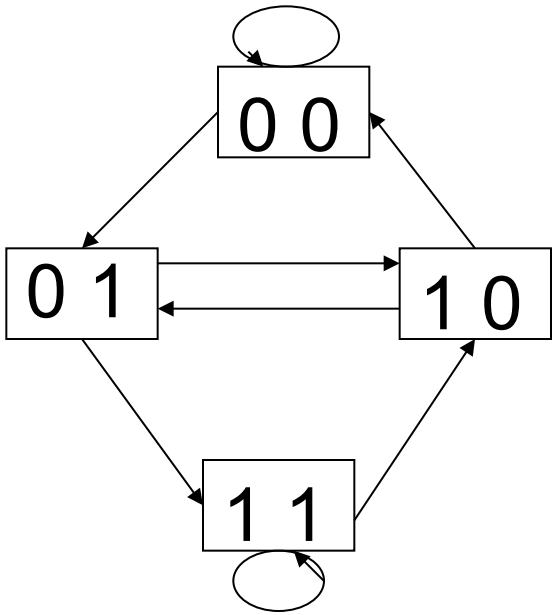
$$(\alpha_0\ \alpha_1\ \cdots\ \alpha_{n-1}) \rightarrow (\alpha_1\ \alpha_2\ \cdots\ \alpha_{n-1}\ 0)$$
$$(\alpha_0\ \alpha_1\ \cdots\ \alpha_{n-1}) \rightarrow (\alpha_1\ \alpha_2\ \cdots\ \alpha_{n-1}\ 1)$$

- Each state has two predecessors

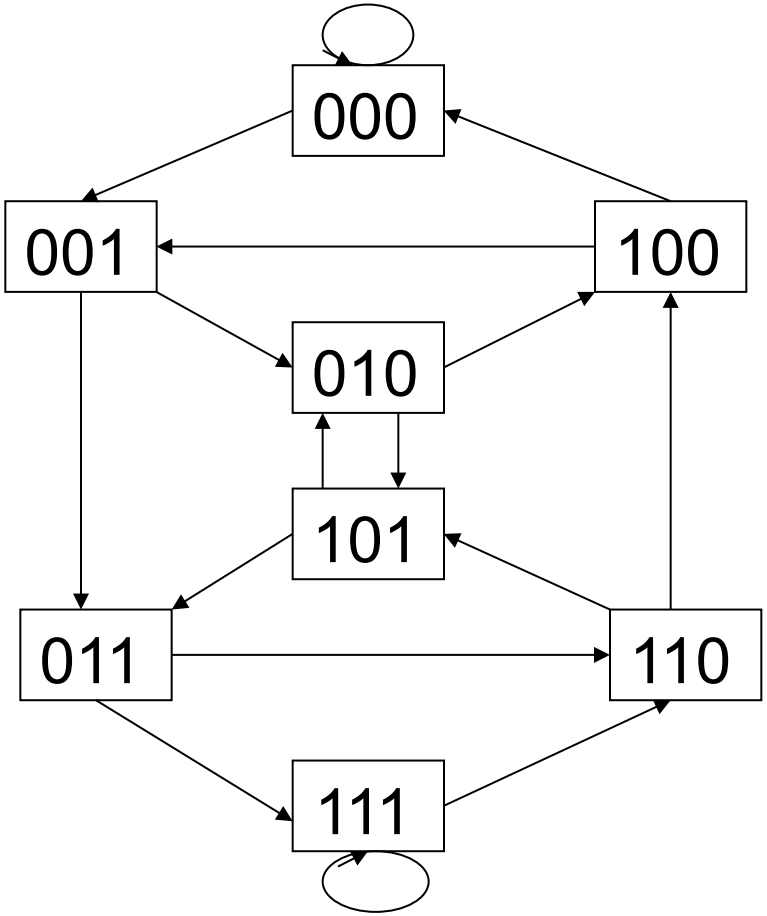
$$(0\ \alpha_1\ \alpha_2\ \cdots\ \alpha_{n-1}) \rightarrow (\alpha_1\ \alpha_2\ \cdots\ \alpha_{n-1}\ 0)$$
$$(1\ \alpha_1\ \alpha_2\ \cdots\ \alpha_{n-1}) \rightarrow (\alpha_1\ \alpha_2\ \cdots\ \alpha_{n-1}\ 1)$$

De Bruijn Graphs (B_2 and B_3)

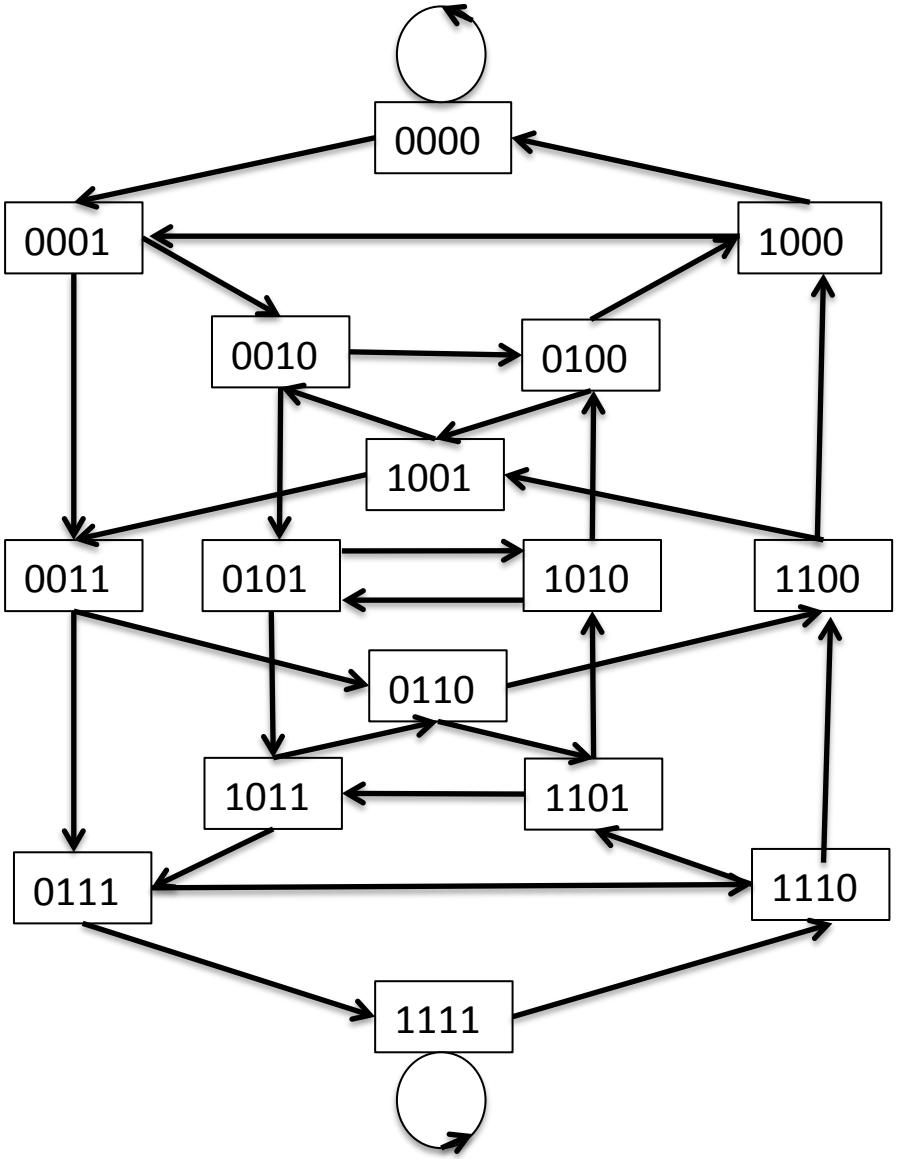
B_2



B_3



De Bruijn graph B_4



Pure Cycling Register (PCR_n)

- Let $f(s_0,s_1,\ldots,s_{n-1}) = s_0$ i.e., $g=0$ (since $f=s_0+g(s_1,\ldots,s_n)$)
  - Weight of truth table of g is 0
  - Cycle structure (PCR_n)

n=3 (0), (1), (001), (011)

n=4 (0), (1), (01), (0001), (0011), (0111)

- Number of cycles of B_n

$$Z(n) = \frac{1}{n}\sum_{d|n}\varphi(d)2^{n/d} \quad (\quad = \quad even \quad number \quad)$$

Pure Cycling Register (PCR_3) : ($f = s_0$)

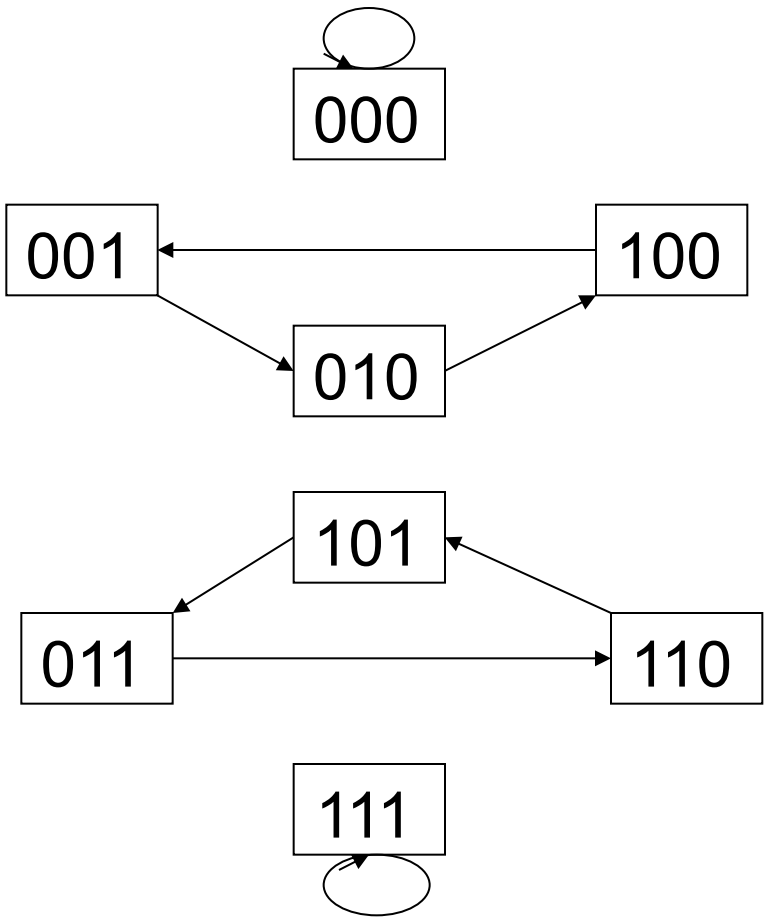
- Decomposition of B_3 for Boolean function $f=s_0$

(0)

(001)

(101)

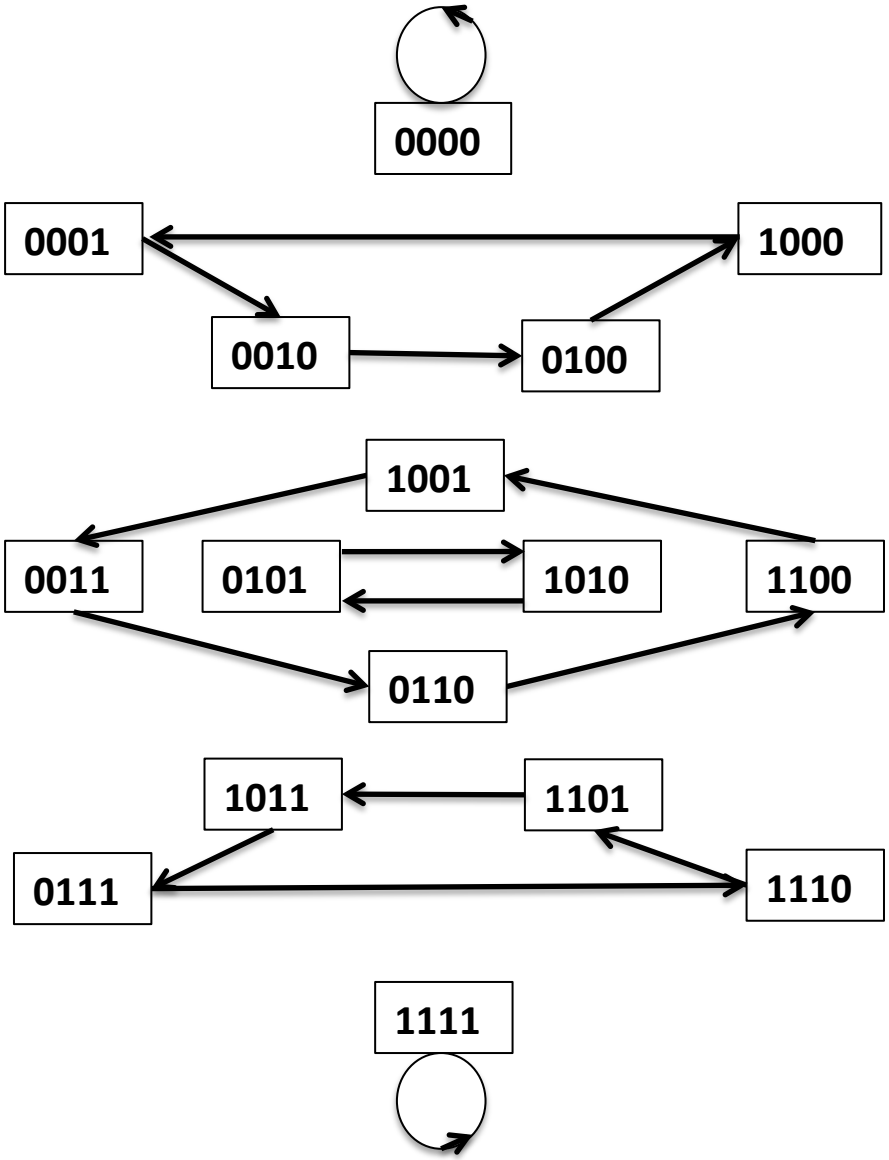
(1)



$f = s_0$

Number of cycles
$Z(3) = 4$

Pure Cycling Register (PCR_4)



(0)

(0001)

(1001)
(01)

(1011)

(1)

Golomb's Conjecture

Golomb's conjecture (1967)

The maximum number of cycles obtained in any decomposition of the de Bruijn graph B_n (for all nonlinear functions f) is Z(n). This occurs for the PCR_n when g=0 (but also in many other cases).

History (approx.)

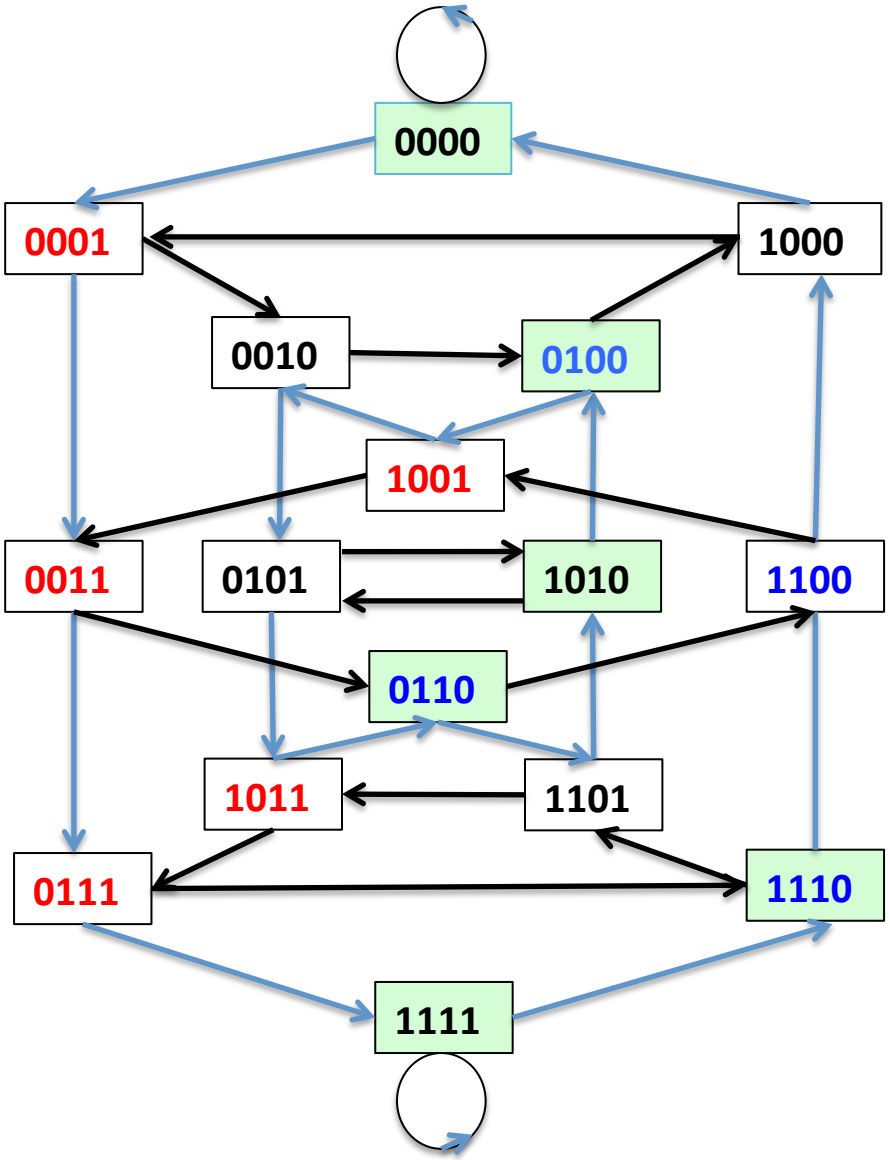
- S. Golomb n=5 / H. Fredricksen n=6, 7 / A. Lempel n=8, 9, 10 / J. Mykkeltveit and Fredriksen n=11,12 ..
- Proved by J. Mykkeltveit (1972), for all n (one year of work to color B_n)

Main idea

Select one node from each cycle in PCR_n (i.e, Z(n) nodes) such that:

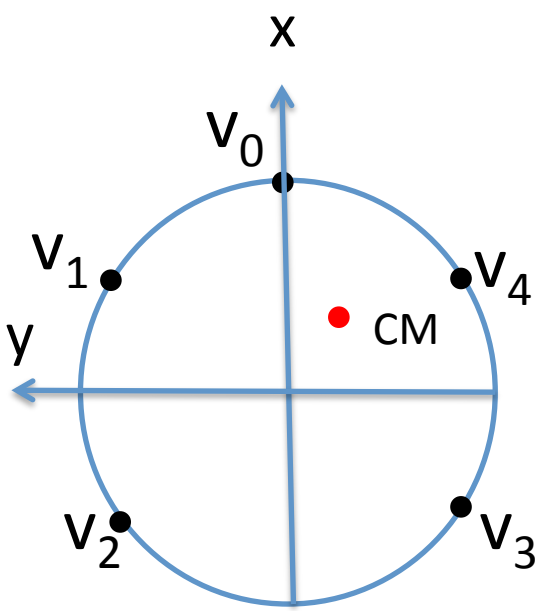
any cycle in B_n contains at least one of these nodes.

Coloring de Bruijn graph B_4



- Any cycle in B_4 contains at least one of the $Z(4)=6$ green colored nodes
- Coloring due to Mykkeltveit
- How to select green color?

CM of a binary n-tuple



Let $V_0=(v_0,v_1,v_2,v_3,v_4)$, (n=5)

Place v_t in coordinate position

$$(x,y)=\left(\cos\frac{2\pi it}{n},\sin\frac{2\pi it}{n}\right)$$

Compute CM=Center of mass

Moment y = $m_{V_0}=\sum_{t=0}^{n-1} v_t \sin\frac{2\pi it}{n}$

Color a vector (v_0, v_1, … , v_{n-1})

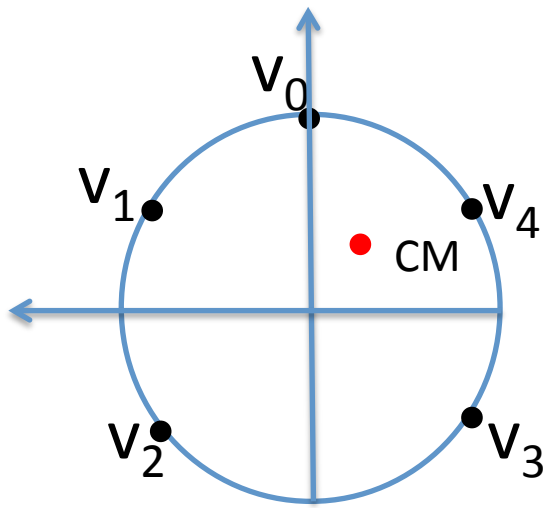
L = If CM on the **left** of the x-axis (y > 0)

I = If CM **on** the x-axis (y = 0)

R = If CM on the **right** of the x-axis (y < 0)

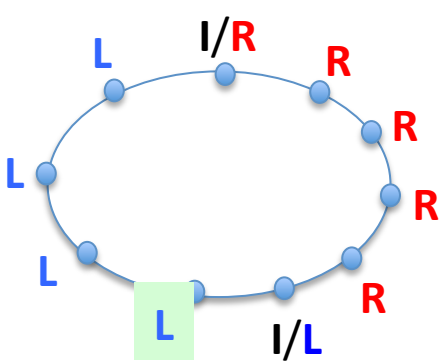
Coloring the PCR_n Cycles

Coloring **L** **I** **R**



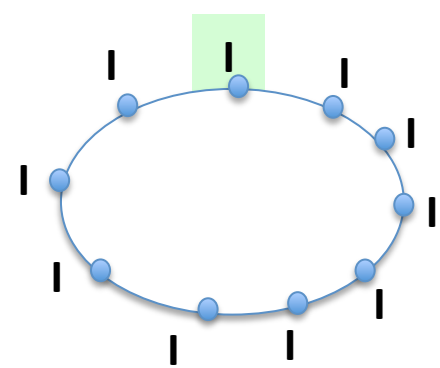
Type 1: (CM not in center of PCR cycle)

- **Select** unique node **L** with predecessor not **L**)



Type 2: (CM in the center of PCR cycle)

- Select **any** node colored **I**



Remarks-Coloring

- Shifting a node cyclically shifts **CM**
- The two predecessors for a node in B_n have the same color (since they only differ in 0-th coordinate on the x-axis).
- The two successors of a node can not both have color I (since they only differ in position n-1).
- A cycle in PCR_n has either:
  - All nodes colored I
  - One R block and one L block separated by most one I.
- Any cycle $S = (s_0, s_1, \ldots, s_{e-1})$ in B_n has (average moment = 0), i.e. has either:
  - All nodes colored I
  - At least one R and one L separated by most one I.

Colors on a cycle

Lemma 1

Let $(s_0,s_1,\ldots,s_{e-1})$ be a cycle of length e on B_n. The nodes (n-tuples) of the cycles are $\mathbf{S}_t=(s_t,s_{t+1},\ldots,s_{t+n-1})$, t=0,1,…,e-1.

Then either

- All nodes on the cycle have the color **I**
- Cycle contains at least one **R** and one **L**

Proof. This follows since the sum of the moments of the y-coordinates on the nodes on a cycle is

$$\sum_{t=0}^{e-1} m_{S_t} = \sum_{t=0}^{e-1}\sum_{t'=0}^{n-1} s_{t+t'} \sin\frac{2\pi t'}{n} = \sum_{t=0}^{e-1} s_t \sum_{t'=0}^{n-1} \sin\frac{2\pi t'}{n} = 0$$

Proof of Golomb's conjecture

Theorem (Mykkeltveit)

No decomposition of the de Bruijn graph B_n for any nonsingular Boolean function f can give more cycles than the PCR_n.

Proof. Select $Z(n)$ nodes one node from each PCR_n cycle.

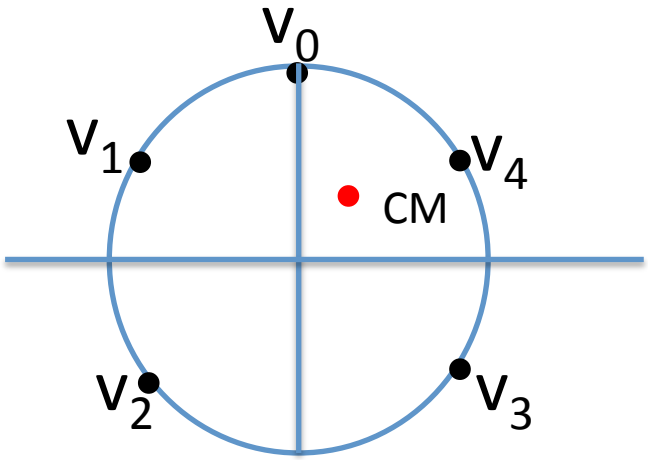
(1) If CM in **center** select arbitrary node on cycle.

(2) If CM **not in center** select first **L** with predecessor not **L**.

Then any cycle in any decomposition will contain at least one of these $Z(n)$ nodes.

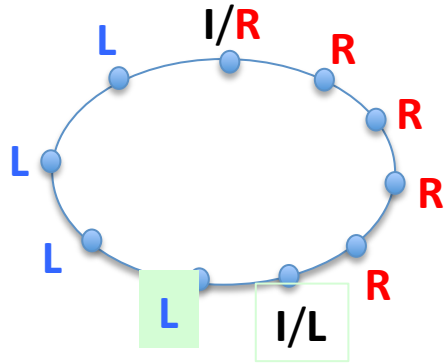
Overview - Proof

Coloring **L** **I** **R**

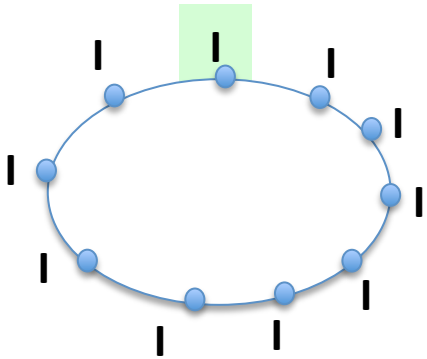


PCR - Cycles

Nodes L and R

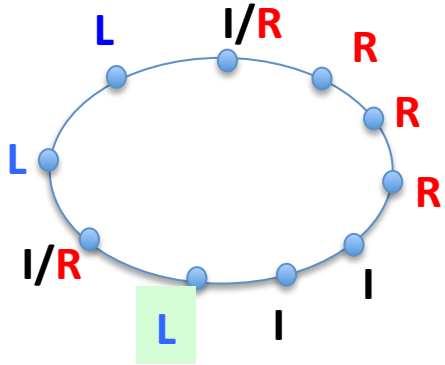


All nodes I

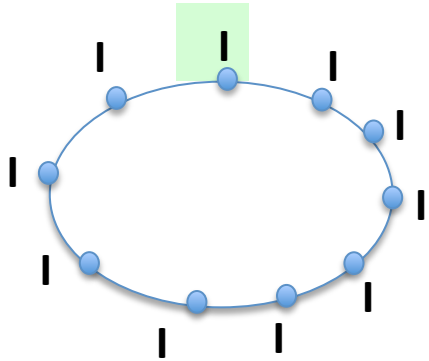


Arbitrary Cycle

Select node on cycle with color **L** and predecessor not **L** is the first **L** in a block of **L**'s on PCR



A cycle with only **I**'s is a PCR cycle with (CM in center)



Cycle Join Algorithm – Joining Cycles

Definition

(α, α^*) is a conjugate pair iff $\alpha + \alpha^* = (1,0,\ldots,0)$

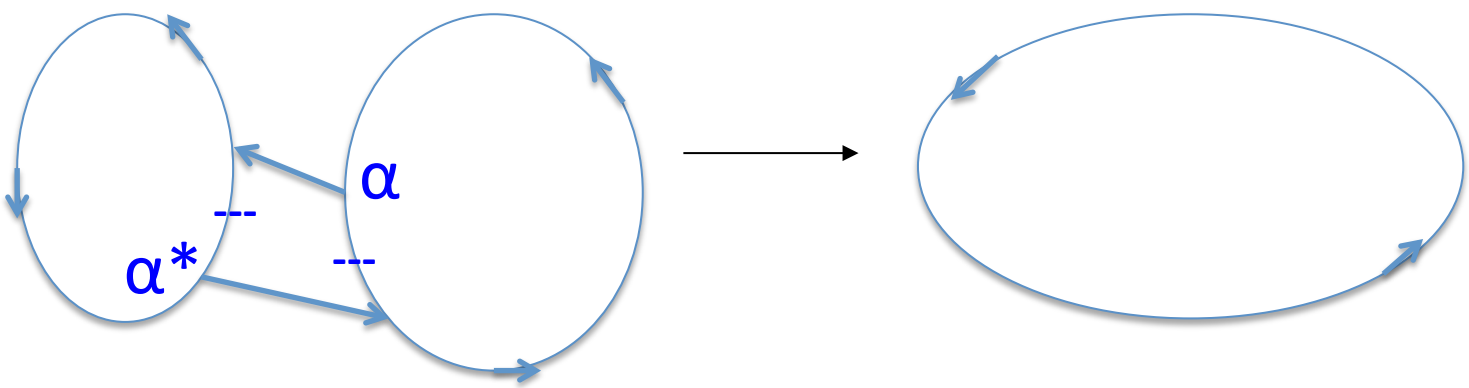
A conjugate pair have the same two possible successors

$(0,\ldots) = \alpha$ $\quad$ $(\ldots,0)$

$(1,\ldots) = \alpha^*$ $\quad$ $(\ldots,1)$

Exchanging successors of (α, α^*) changes $g(\ldots)$ for only one value

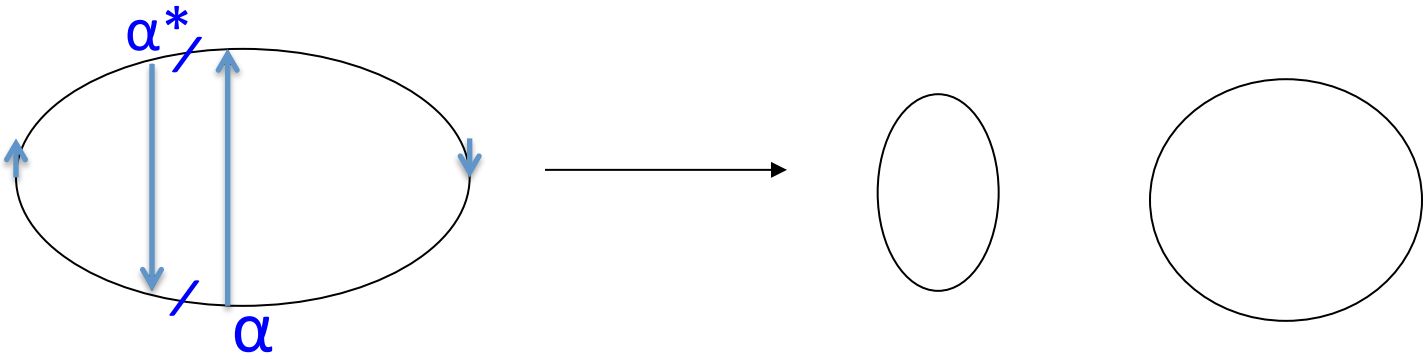
Joining two cycles–Change successors of α and α^* on different cycles



Splitting a Cycle

Splitting a cycle

- Exchanging the successors of a conjugate pair (α, α*) on the same cycle
- This change parity of truth table g by one (and also changes parity of number of cycles by one)



Parity of Number of Cycles

Theorem

The number of cycles which B_n is composed into has the same parity as the weight of the truth table of g

Proof:

The function $f=x_0+g$ where g=0 gives Z(n) (even) cycles

- Any other nonlinear function f can be obtained by changing truth table bit by bit.
- Each change of truth table of g changes the number of cycles by one **and** the weight of g by 1

Hence, parity stays invariant between cycles and weight

DeBruijn sequences (Necc. conditions)

Theorem

(1) To obtain a deBruijn sequence then f uses all n variables

(2) The truth table of g ($f=s_0+g$) must have odd weight (at least $Z(n)-1$)

Proof: Follows since otherwise truth table has even weight and can not generate a de Bruijn sequence

De Bruijn sequences from m-sequences

- Change longest run in m-sequence by appending an extra 0. The result is a deBruijn sequence
- Example: 0000100110101111
- This de Bruijn sequence is ”almost linear”
- However, linear complexity is as large as possible for deBruijn sequences
- This is a prime example that linear complexity is no guarantee for security
- Bounds on the linear complexity of de Bruijn sequences is studied (Chan, Games 1980s)

Periods on NLFSRs

1. General
2. Kjeldsen's method
3. Mykkeltveit and AN-codes

Period of Nonlinear Shift Registers

- Hard problem in general
- Rather few general results on the period
- Some nontrivial results known in the case when $g(x_1,\ldots,x_{n-1})$ is a symmetric polynomial

  (Kjeldsen, Søreng from the 1970-80s)
- Proofs are in general very technical and hard to read and new simpler methods are needed to progress
- Mykkeltveit (1979) used arithmetic codes to study periods of nonlinear shift registers
- Classification of de Bruijn sequences

  (Fredricksen 1982, Hauge and Mykkeltveit 1990's)

Kjeldsen's Mapping (1)

δ: $x_i \rightarrow x_{i+1}$ for *i=0,1,…,n-2*

$x_{n-1} \rightarrow x_n = x_0 + g(x_1,\ldots,x_{n-1})$

This algebra homomorphism leads to a sequence of polynomials in the polynomial ring $F[x_0,x_1,\ldots,x_{n-1}]/(x_0^2+1, \ldots ,x_{n-1}^2+1)$

$$(x_0,x_1,\ldots,x_{n-1},x_n,x_{n+1}, \ldots\ldots,x_{t+n}= x_t + g(x_{t+1},\ldots,x_{t+n-1}), \ldots.)$$

Definition

The period of δ is the smallest integer p such that $\delta^p = id$.

(= smallest period of x_0)

Theorem

All sequences in $\Omega(f)$ (generated by $s_{t+n}= s_t+ g(s_{t+1},\ldots,s_{t+n-1})$)

have period dividing p and at least one sequence has least period p.

Kjeldsen's Mapping (2)

δ: $x_i \rightarrow x_{i+1}$ for *i=0,1,…,n-2*

$x_{n-1} \rightarrow x_n = x_0 + g(x_1,\ldots,x_{n-1})$

Let $h(x_0,\ldots,x_{n-1}) = h_1(x_0,\ldots,x_{n-2}) + x_{n-1}h_2(x_0,\ldots,x_{n-2})$

Then

$$\begin{aligned}\delta(h) &= h_1(x_1,\ldots,x_{n-1}) + (x_0+g)h_2(x_1,\ldots,x_{n-1})\\ &= h_1(x_1,\ldots,x_{n-1}) + x_0h_2(x_1,\ldots,x_{n-1}) + gh_2(x_1,\ldots,x_{n-1})\\ &= h(x_1,\ldots,x_{n-1},x_0) + g(x_1,\ldots,x_{n-1})\, h_2(x_1,\ldots,x_{n-1})\\ &= h(\sigma(x_0,\ldots,x_{n-1})) + g(x_1,\ldots,x_{n-1})h_2(x_1,\ldots,x_{n-1})\end{aligned}$$

where σ is cyclic shift of n-tuples. Hence, defining $g_2 = g^*$

$$\delta(g) = \sigma(g(x_0,\ldots,x_{n-1})) + g^*g$$

Symmetric Feedback Polynomials

Let S_j be elementary symmetric polynomial of degree j in $n-1$ variables

Theorem (Kjeldsen)

If $g(x_1,\ldots,x_{n-1}) = \sum_{k=0}^{(n-2/2} a_k S_{2k+1}(x_1,\ldots,x_{n-1})$, $a_k \in \{0,1\}$, $g \neq 0$, S_1,

then the minimal period of δ is $n(n+1)$.

Proof sketch: It follows that due to symmetry in g we can derive the condition

$(\sigma^j g)^* g = 0$ if $j = -1 \pmod n$ (since independent of x_{n-1})

$= g$ if $j \neq -1 \pmod n$ (periodic in j and symmetry)

Hence, $\delta(g) = \sigma(g(x_0,\ldots,x_{n-1})) + gg^* = \sigma(g(x_0,\ldots,x_{n-1})) + g$ where

$$g^*(x_1,\ldots,x_{n-1}) = \sum_{k=0}^{(n-2/2} a_k S_{2k}(x_2,\ldots,x_{n-1})$$

Proof Remarks

Note $\delta(g) = \sigma(g(x_0,\ldots,x_{n-1})) + g*g = \sigma(g) + g$

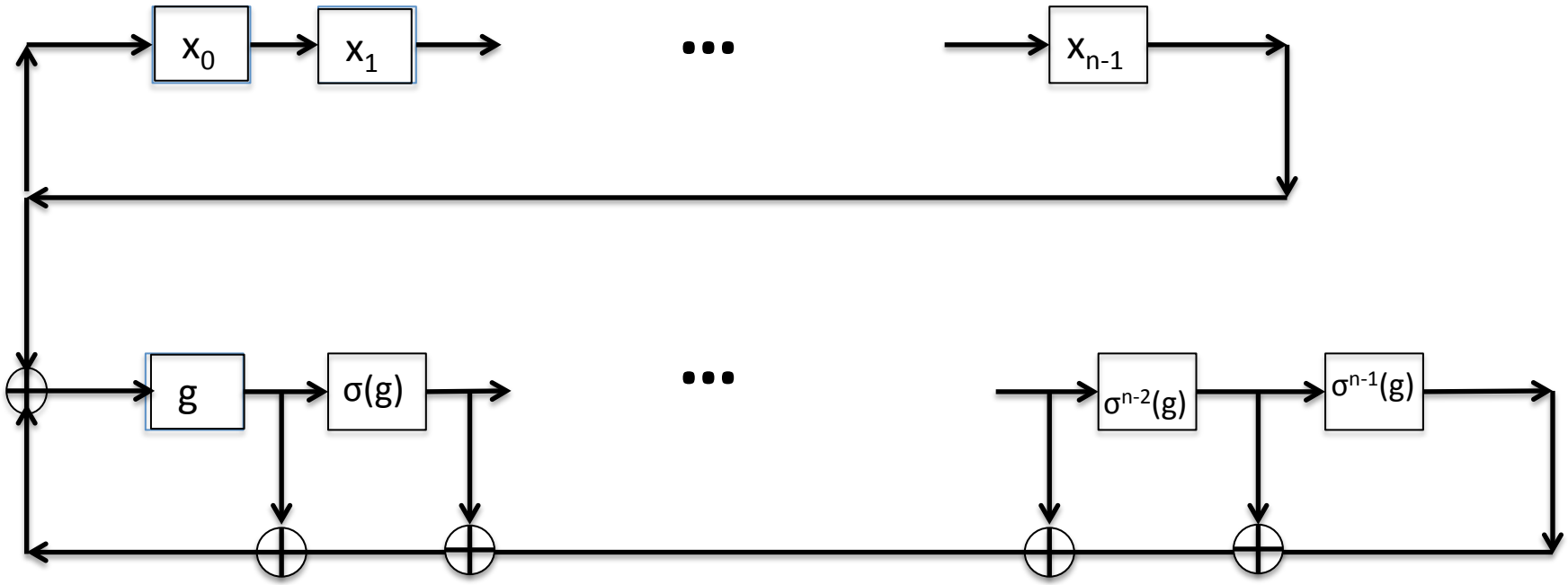
Therefore

$$\delta(\sigma(g)) = \sigma^2(g(x_0,\ldots,x_{n-1})) + (\sigma g)*g$$
$$= \sigma^2(g(x_0,\ldots,x_{n-1})) + g*g = \sigma^2(g) + g$$

and in general

$$\delta(\sigma^{n-1}(g)) = g \quad \text{for } j = -1 \bmod n$$
$$\delta(\sigma^{j}(g)) = \sigma^{j+1}(g) + g \quad \text{for } j \neq -1 \bmod n$$

Kjeldsen's Method – II



- "Linear register" of period n(n+1)
- Characteristic polynomial $(x^n+1)(x^{n+1}+1)/(x+1)$
- Provided some suitable conditions on g (like gg* = g etc.)
- Many symmetric polynomials g satisfy conditions
- Lead to controllable periods n(n+1)
- Even though "small" period the was important idea

Period of Nonlinear Register and Coding

Theorem

Let C be a cyclic code (not necessarily linear) with $d_{min} \geq 3$. Define $f = x_0 + g$ where

$g(x_1,\ldots x_{n-1})=1$ iff $(0,x_1+1,\ldots x_{n-1}+1) \varepsilon C$ or $(0,x_1+1,\ldots x_{n-1}+1) \varepsilon C$.

Then all sequences in $\Omega(f)$ have periods dividing $n(n+1)$.

Proof:

Follows since also in this case

$$\sigma^j(g)g^* = 0 \text{ if } j = -1 \pmod n$$
$$= g \text{ if } j \neq -1 \pmod n$$

AN-Codes and Period of NLFSRs

An AN-Code is an arithmetic code is a code with codewords

$$C = \{AN \pmod{2^n-1} : N=0,1,\ldots,B-1\}$$

where $AB=2^n-1$.

The codewords AN can be represented binary $(a_0,a_1,\ldots,a_{n-1})$

$$a_i = (N\cdot 2^i \pmod{B}) \pmod{2}$$

The codewords have period dividing n and can be defined via NLFSRs

Mykkeltveit (1977) determined the corresponding NLFSRs for the codewords in the AN-code for several values of A and thus their periods.

Algebraic Methods for NLFSRs

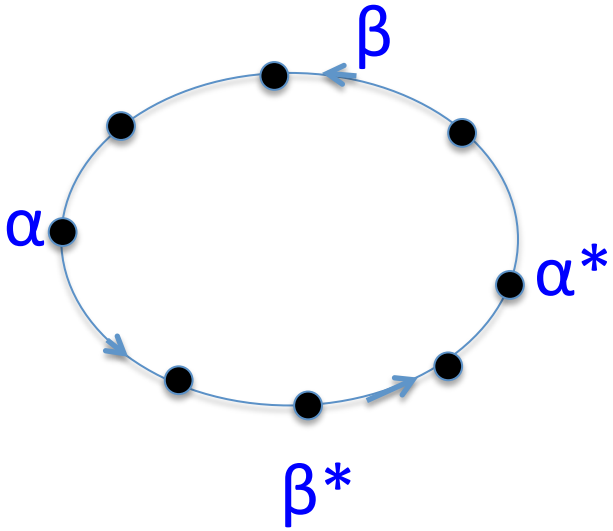
- Cross-join pairs on a cycle
  - Two conjugate pairs (α,α^*) and (β, β^*) on a cycle such that interchanging the successors of each of these pairs give the same number of cycles ("split and join").
  - The number of cross-join pairs were conjectured for m-sequences by Kim et. al. in 1990 and solved Helleseth and Kløve using simple connections with finite field

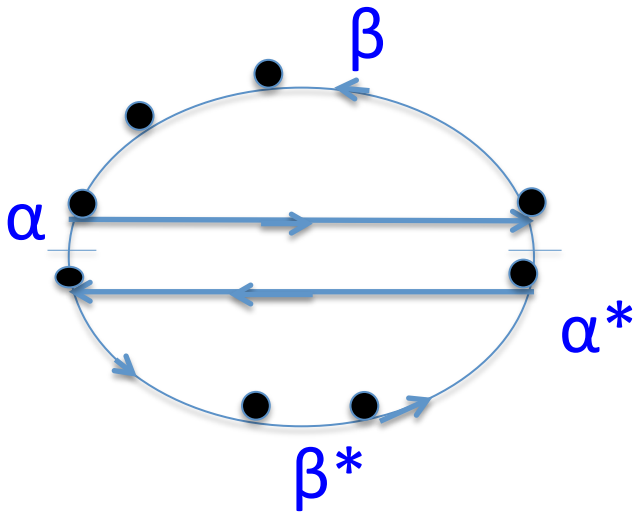
- Cyclotomy and the number of conjugate pairs from irreducible cyclic codes
  - An irreducible cyclic code of period $e|2^n-1$ decomposes B_n into $E=(2^n-1)/e$ disjoint cycles.
  - Using a special mapping between nodes in B_n reduces problem of finding conjugate pairs on the E cycles to
  - This gives estimate of number of de Bruijn sequences that can be constructed by joining the cycles from the irreducible code

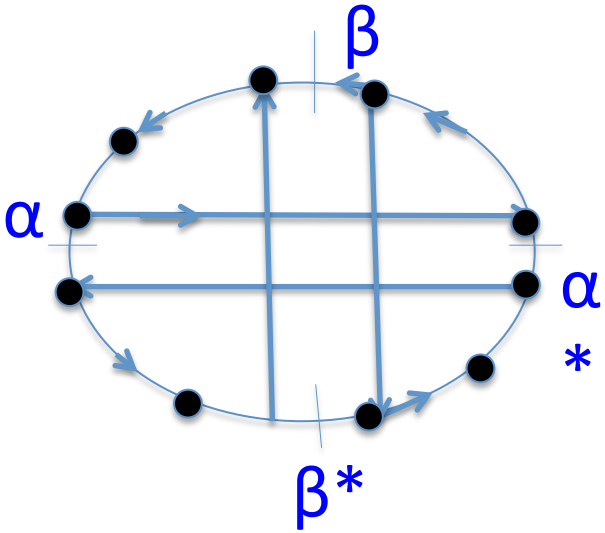
Cross Join Pairs

(α,α^*) and (β,β^*) conjugate pairs

$\alpha + \alpha^* = \beta + \beta^* = (1,0,\ldots,0)$







Cross Join Pairs on m-sequences

Given an m-sequence

1. Split the cycle into two cycles using a conjugate pair (α,α^*) on m-sequence
2. Join the two cycles into one cycle using a new conjugate pair (β, β^*) (on the two new cycles)

The pair (α,β) is called a cross-join pair

Theorem (Helleseth and Kløve)

The number of cross-join pairs on an m-sequence is

$$N = (2^{n-1}-1)(2^{n-1}-2)/6$$

Mapping

Mapping ϕ between F_2^n and F_2^n

Example: $\Psi^4 + \Psi^3 + 1 = o$

	1	Ψ	Ψ^2	Ψ^3
1	1	0	0	0
Ψ	0	1	0	0
Ψ^2	0	0	1	0
Ψ^3	0	0	0	1
Ψ^4	1	0	0	1
Ψ^5	1	1	0	1
Ψ^6	1	1	1	1
Ψ^7	1	1	1	0
Ψ^8	0	1	1	1
Ψ^9	1	0	1	0
Ψ^{10}	0	1	0	1
Ψ^{11}	1	0	1	1
Ψ^{12}	1	1	0	0
Ψ^{13}	0	1	1	0
Ψ^{14}	0	0	1	1

Let s_t be the first coordinate sequences.

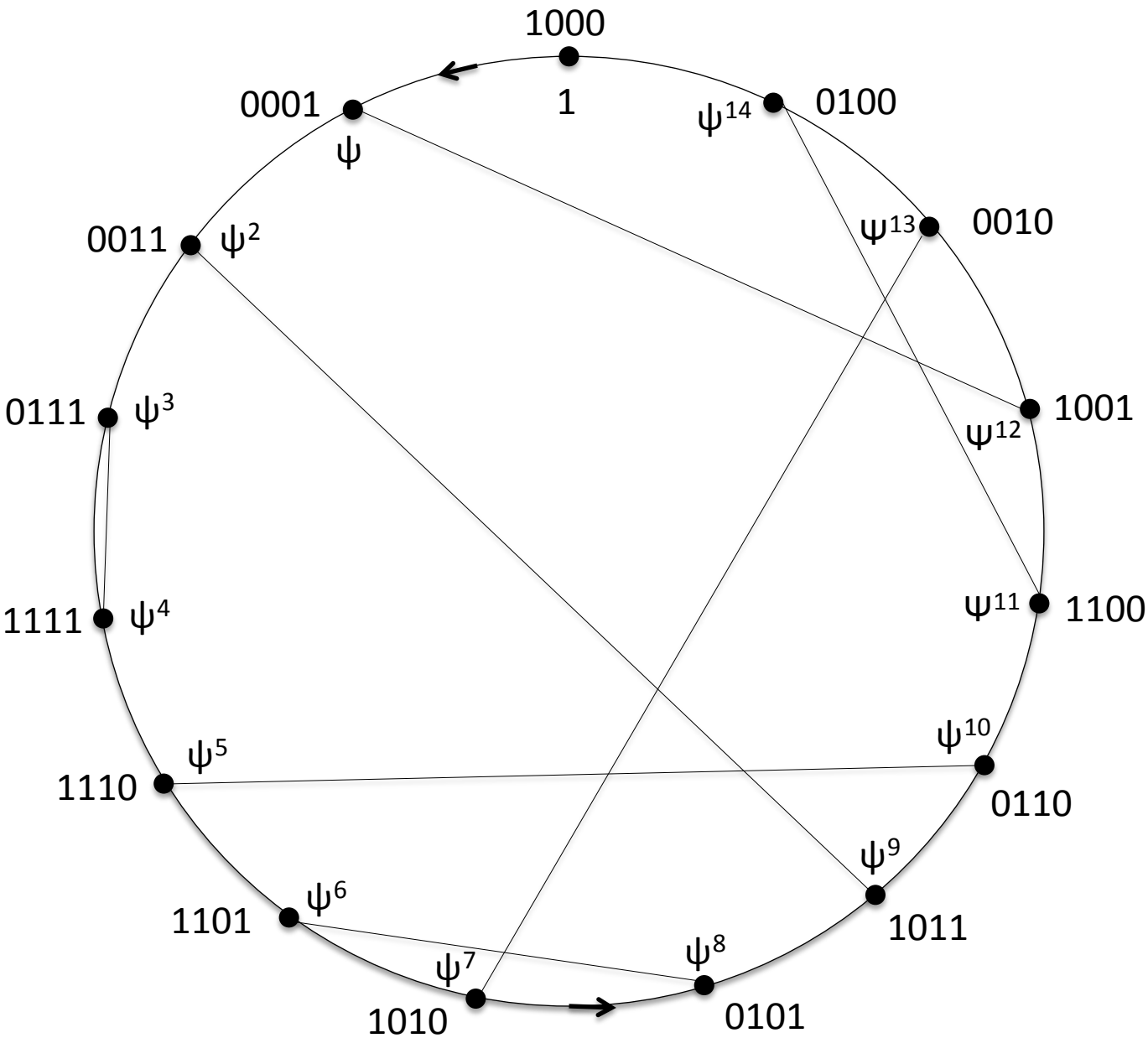
Then

$$\Phi(0) = (0, 0, \ldots, 0)$$

$$\Phi(\Psi^t) = (s_t, s_{t+1}, \ldots, s_{t+n-1})$$

Conjugate pairs (x, x^*) correspond to elements with $x + x^* = 1$

Cross-join pairs corresponds to equivalence classes of intersecting chords


1000
1
0001
ψ
0100
ψ14
0010
ψ13
0011
ψ2
1001
ψ12
0111
ψ3
1100
ψ11
1111
ψ4
ψ10
0110
ψ5
1110
ψ9
1011
ψ6
1101
ψ8
0101
ψ7
1010

Number of Cross Join Pairs

- One-to-one correspondence between cross join pairs and equivalence classes of subsets $\{\theta_1, \theta_2, \theta_3, \theta_4\}$ with $\theta_1 + \theta_2 + \theta_3 + \theta_4 = 0$ (wlog $\{\theta_1, \theta_3\}$ and $\{\theta_2, \theta_4\}$ are intersecting
- Two sets are equivalent iff $\theta\{\theta_1, \theta_2, \theta_3, \theta_4\} = \{\theta_1, \theta_2, \theta_3, \theta_4\}$
- The number of distinct subsets are

$$(2^n - 1)(2^n - 2)/24$$

- Each equivalence class contains exactly one cross-join pair. Thus dividing by 2^n-1 gives the number of cross join pairs
- The cross join pair corresponds to the unique θ with

$$\theta\theta_1 + \theta\theta_3 = \theta\,\theta_2 + \theta\theta_4 = 1$$

Cyclotomy and Cross Join pairs

Let C be irreducible cyclic code of period $e = (2^n - 1)/E$

$$C=\{ c_a \mid (c_a)_x = Tr(ax^E),\ a\varepsilon GF(2^n)\}$$

Code consists of E cycles of period e.

The cyclotomic classes

$$C_i = \{\Psi^{tj+i} : 0 \le t < (2^n-1)/E)\} \text{ for } i=0,1,\ldots,E-1.$$

The cyclotomic numbers (i,j) of order E is the number of solutions

$$z_i + 1 = z_j$$

where z_i, z_j belong C_i and C_j respectively.

Mapping of Cycles

Similar mapping as for cross-join pairs

Nodes in cycle i can be represented by

$$\Psi^i \beta^t \text{ , } t=0,1,\ldots,e-1$$

where β is zero of irreducible (pairty-check) polynomial of the code.

The number of conjugate pairs between cycle i and j is the number of solutions $z_i+1=z_j$ which is the cyclotomic number.

The number of de Bruijn sequences obtained by joining cycles in the irreducible code can be estimated from the "BEST" theorem that gives the number of spanning trees in the Cycle Joining Algorithm (CJA)

Conclusions

- NLFSRs is an important topic with great potential
- Many open problems
  - Period
  - Distribution
  - Construction of sequence families
- Most results are quite old
- New ideas are needed to solve the challenges in the analysis of nonlinear generated sequences