

Design of Pseudo-random Sequence Generators (PRSG)

Lecture 1:

- Overview of PRSG
- Linear Feedback Shift Register Sequences
- Randomness Measurements

Lecture 2:

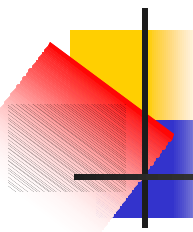
- Design of PRSGs Towards Large Linear Span
- Applications in Stream Cipher Design: A5 and w7

Developments of Pseudo-random Sequences



Three Periods of Research for Pseudo-Random Sequences

- Period of pre-application (before 1948)
- Golden period of m-sequences (1948-1969)
- Period of non-linear generators (1969 - present)



Golden period of m -sequences (1948-1969)

- Shannon's Result (1948): One-time-pad is unbreakable
- Berlekamp-Massey algorithm attack (1969)
 - LFSR generates m -sequences
 - = Maximal length sequences
 - = Pseudo-noise sequences
 - = PSG (1948-1969)

0. Overview of PRSG (Cont.)



Period of Non-Linear Generators

- Design towards 2-Level Auto-Correlation and Low Correlation

- Design towards Large Linear Span



LFSR as Basic Blocks

Implemented as

- Key Stream Generators in Stream cipher Models in Wireless Environment
- Functions in Block Ciphers
- Session Key Generators
- Pseudo-random Number Generators in Digital Signature Standard (DSS), ECDSS, etc.
- Digital Water-mark

....



1. Linear Feedback Shift Registers (LFSR)

- **Feedback Shift Registers (FSR)**
- **Characteristic Polynomials and Periods of LFSR**

✦ 1.1 Feedback Shift Registers (FSR)

A. Basic Concepts and Examples: Binary Case

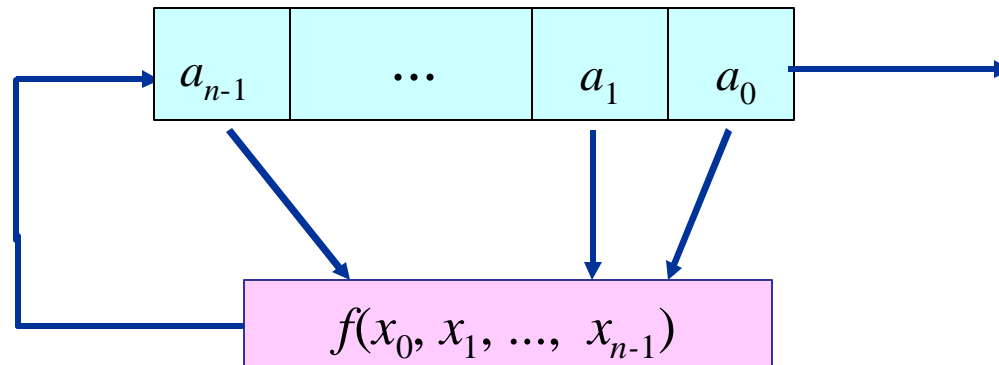


Fig 1. A Block Diagram of an FSR



Three Components

(1) n -stage shift register : n 2-state storage units

(2) Initial state $(a_0, a_1, \dots, a_{n-1})$

(3) Feedback function : $f(x_0, \dots, x_n)$ is a boolean function in n variables:

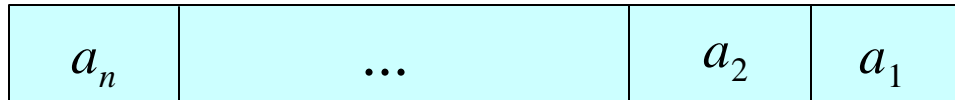
$$f(x_0, \dots, x_{n-1}) = \sum_{\{i_1, \dots, i_s\} \subset \{0, \dots, n-1\}} a_{i_1 \dots i_s} x_{i_1} \cdots x_{i_s}, \quad a_{i_1 \dots i_s} \in \{0, 1\}$$

There are 2^{2^n} Boolean functions in n variables.

How does it work?

At each clock pulse: the state of each memory stage is shifted to the next stage in line, *i.e.*, there is a transition from one state to next.

For example, the next state of Fig. 1 is



where

$$a_n = f(a_0, a_1, \dots, a_{n-1})$$

and the device outputs one bit a_0 .

So, we have a sequence

$$\{a_i\} = a_0, a_1, \dots, a_{n-1}, a_n, \dots$$

where the recursive relation is given by

$$a_{n+k} = f(a_k, a_{k+1}, \dots, a_{k+n-1}) \text{ for } k = 0, 1, \dots$$



The sequence $\{a_i\}$ is said to be an FSR sequence and a vector

$$\mathbf{s}_i = (a_i, a_{i+1}, \dots, a_{i+n-1}), i = 0, 1, \dots,$$

is called the *ith state* of the FSR or the sequence.

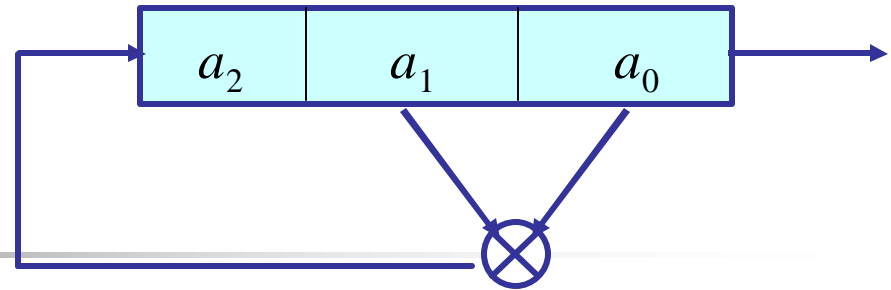
If $f(x_0, x_1, \dots, x_{n-1})$ is linear, *i.e.*,

$$f(x_0, x_1, \dots, x_{n-1}) = c_0x_0 + c_1x_1 + \dots + c_{n-1}x_{n-1}, c_i \in \{0,1\},$$

then $\underline{a} = \{a_i\}$ is called an LFSR sequence. Otherwise, an nonlinear FSR (NLFSR) sequence.

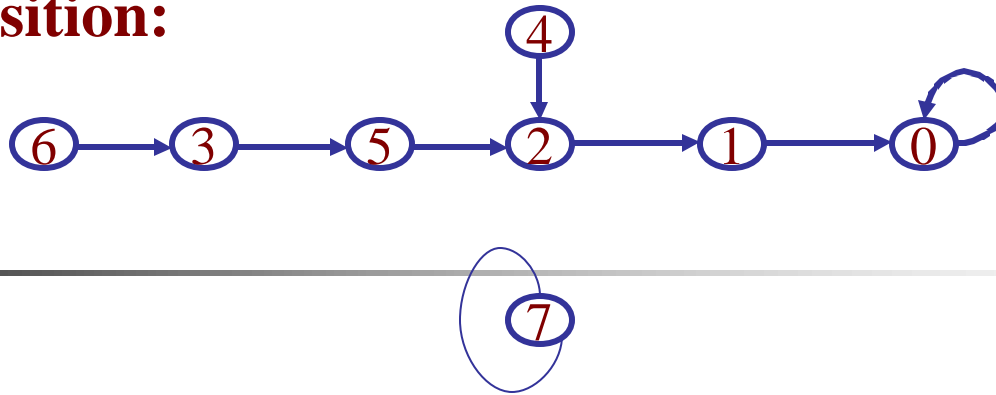
Example 1. A 3-stage NFSR with a feedback function

$$f(x_0, x_1, x_2) = x_0 x_1$$



	Current state	Next state
	$x_2 \ x_1 \ x_0$	$x_2 \ x_1 \ x_0$
0	0 0 0	0 0 0
1	0 0 1	0 0 0
2	0 1 0	0 0 1
3	0 1 1	1 0 1
4	1 0 0	0 1 0
5	1 0 1	0 1 0
6	1 1 0	0 1 1
7	1 1 1	1 1 1

State transition:



Output sequences for different initial states:

100000... $(a_0, a_1, a_2) = (1, 0, 0)$

01101000... $(a_0, a_1, a_2) = (0, 1, 1)$

001000... $(a_0, a_1, a_2) = (0, 0, 1)$

111111... $(a_0, a_1, a_2) = (1, 1, 1)$

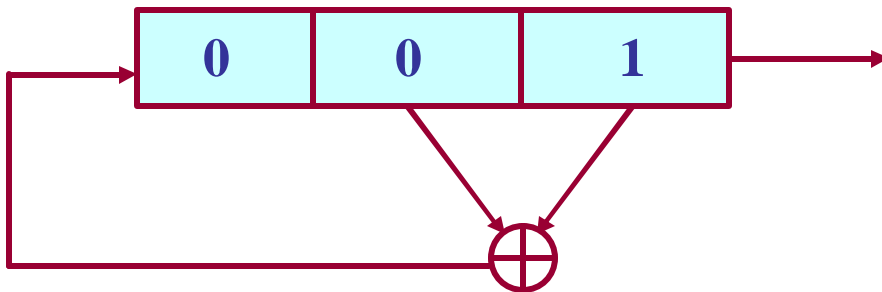
Recursive relation:

$$a_{3+k} = a_k a_{1+k}, \quad k = 0, 1, \dots$$

Example 2. A 3-stage LFSR with a feedback function

$$f(x_0, x_1, x_2) = x_0 + x_1$$

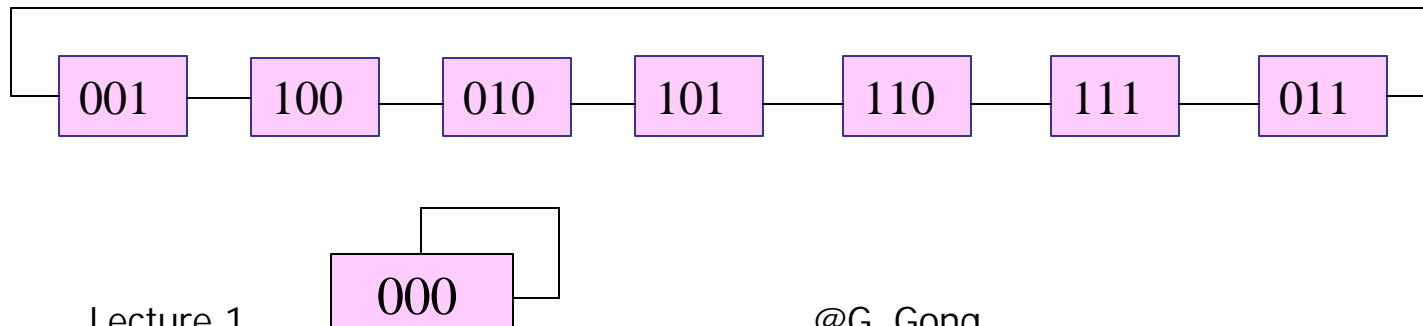
(1) Implementation

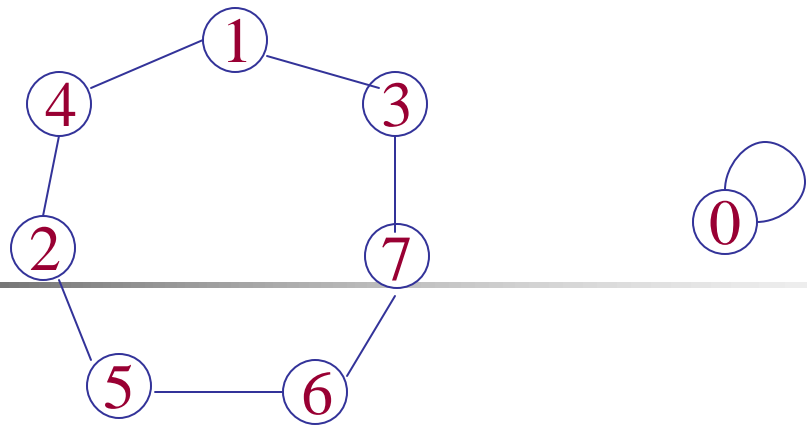
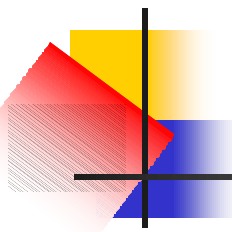


(2) Recursive relation:

$$a_{3+k} = a_{1+k} + a_k, k = 0, 1, \dots$$

(3) State Diagram





(4) Outputs with different initial states:

Initial state: (a_0, a_1, a_2)

Output sequence:

$(1, 0, 0)$:

10010111001011...

$(1, 1, 1)$:

11100101110010...

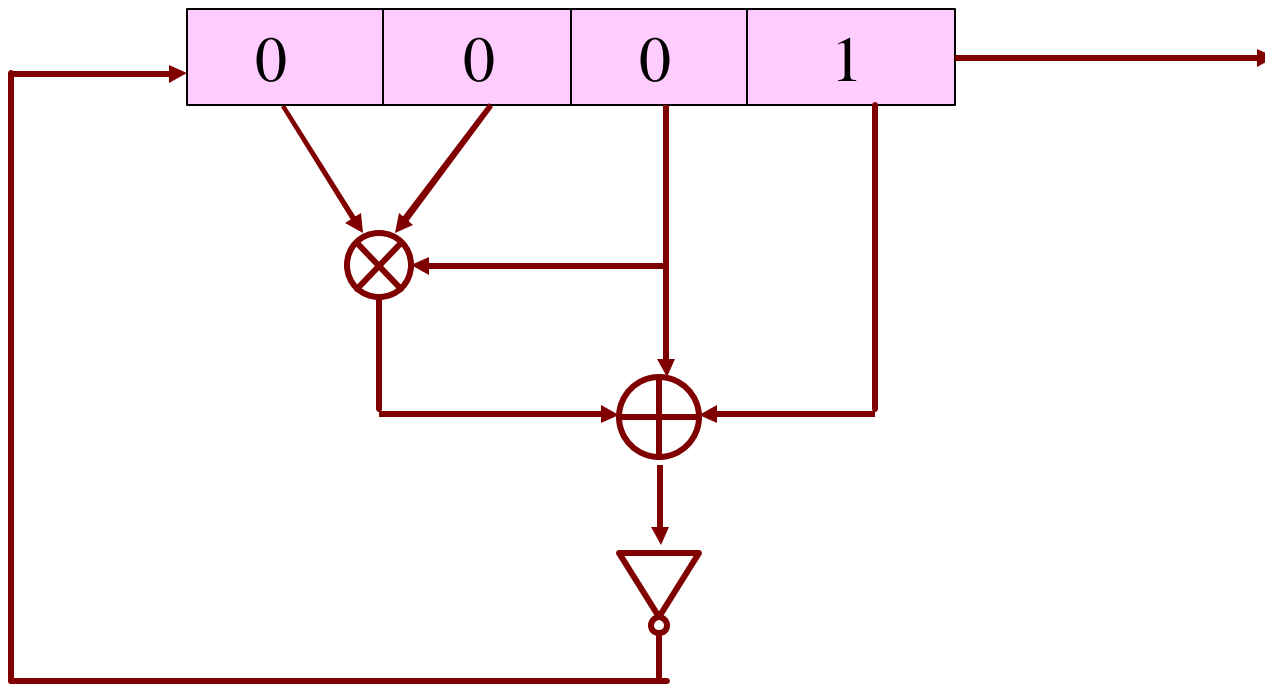
$(0, 0, 0)$:

0000000...

Example 3. Let the feedback function be given by

$$f(x_0, x_1, x_2, x_3) = 1 + x_0 + x_1 + x_1x_2x_3$$

A 4-stage FSR:



Truth Table

$$g(x_0, x_1, x_2, x_3) = x_0 + x_1 + x_1x_2x_3 \text{ and } f = g + 1$$

x_0	x_1	x_2	x_3	g	f	x_0	x_1	x_2	x_3	g	f
0	0	0	0	0	1	0	0	0	0	1	1
1	0	0	0	1	0	1	0	0	1	1	0
0	1	0	0	1	0	0	1	0	1	1	0
1	1	0	0	0	1	1	1	0	1	0	1
0	0	1	0	0	1	0	0	1	1	0	1
1	0	1	0	1	0	1	0	1	1	1	0
0	1	1	0	1	0	0	1	1	1	0	1
1	1	1	0	0	1	1	1	1	1	1	0

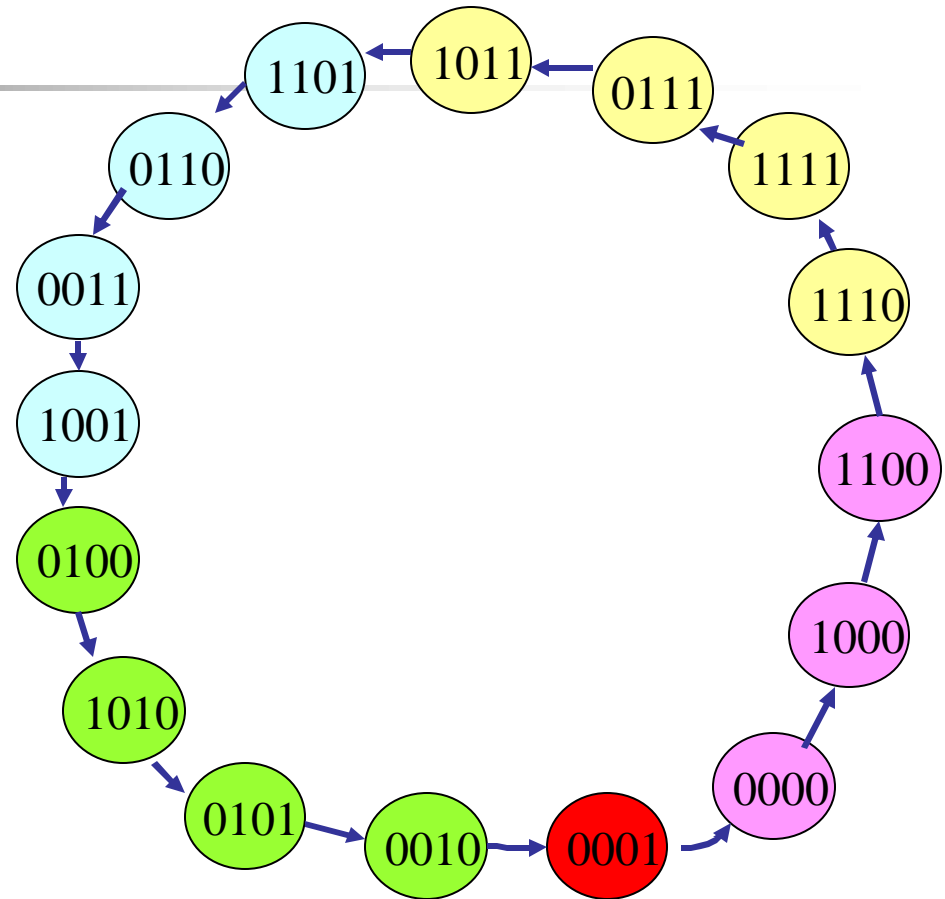
The output sequence:

1000 011110101100

1000011110101100...

which has period 16.

A de Bruijn sequence is an output sequence of n -stage NLFSR with period 2^n .



State Diagram

B. Formal definition of q -ary FSR sequences

An Abstract model: Let

- $F = GF(q)$, a finite field of order q , where q is a prime or a power of prime.
- $f(x_0, x_1, \dots, x_{n-1})$ be a function in n variables defined as

$$f(x_0, \dots, x_{n-1}) = \sum_{\{i_1, \dots, i_s\} \subset \{0, \dots, n-1\}} c_{i_1 \dots i_s} x_{i_1}^{e_{i_1}} \cdots x_{i_s}^{e_{i_s}}, \quad c_{i_1 \dots i_s} \in F.$$

where e_{i_j} are positive integers with $1 \leq e_{i_j} < q$, $0 \leq s \leq n$.

- $\underline{\mathbf{a}} = \{a_i\}$, $a_i \in F$ whose elements are given by

$$a_{n+k} = f(a_k, a_{k+1}, \dots, a_{k+n-1}), \quad k = 0, 1, \dots$$

Then

- $\underline{\mathbf{a}}$ is said to be a q -ary FSR sequence over F ,
- $(a_0, a_1, \dots, a_{n-1})$ is called an initial state of $\underline{\mathbf{a}}$
- $f(x_0, x_1, \dots, x_{n-1})$ is called the feedback function of $\underline{\mathbf{a}}$.

If an feedback function $f(x_0, \dots, x_{n-1})$ is linear, *i.e.*,

$$f(x_0, x_1, \dots, x_{n-1}) = c_0x_0 + c_1x_1 + \dots + c_{n-1}x_{n-1}, \quad c_i \in F,$$

then the recursive relation becomes a linear recursive relation:

$$a_{n+k} = \sum_{i=0}^{n-1} c_i a_{i+k}, \quad k = 0, 1, \dots$$

So an LFSR sequence $\{a_i\}$ is also called a **linear recursive sequence** over F .



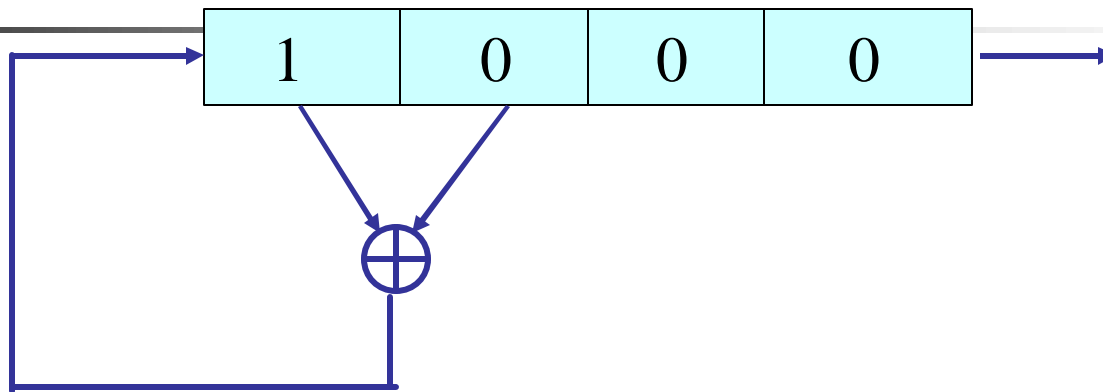
C. Periodic Property

Def. Let $\mathbf{a} = \{a_i\}$, $a_i \in F$. If there exists integer $r > 0$ and $u \geq 0$ such that

$$a_{i+r} = a_i, \text{ for all } i \geq u,$$

then the sequence is said to be **ultimately periodic** and r is called a **period** of the sequence. The smallest integer satisfies by the above identity is called a **least period** of the sequence.

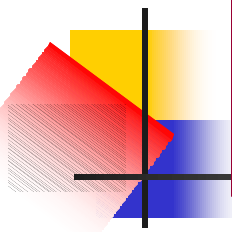
Example 4. Let $F = GF(2)$




4 - stage LFSR with $f(x_0, x_1, x_2, x_3) = x_2 + x_3$.

Output: 00011011011...

which is ultimately periodic with $u = 2$ and $r = 3$.

- 
- Any output sequence of an n -stage FSR over F is ultimately periodic with period $r \leq q^n$. In particular, if $q = 2$, then period $r \leq 2^n$.

- If the feedback function is linear, then any output of the LFSR is ultimately periodic with period $r \leq q^n - 1$. In particular, if $q = 2$, then period $r \leq 2^n - 1$. (Why?)



Question: How can one determine the periodic property of the output sequences of FSR or LFSR? In other words, under which conditions, does the state diagram has no branches?

✦ 1.2 Characteristic Polynomials and Periods of LFSR

Let \mathbf{a} be a LFSR sequence with a feedback function

$$g(x_0, x_1, \dots, x_{n-1}) = \sum_{i=0}^{n-1} c_i x_i, \text{ then } a_{n+k} = \sum_{i=0}^{n-1} c_i a_{i+k}, \quad k = 0, 1, \dots$$

$$\text{Let } f(x) = x^n - c_{n-1}x^{n-1} - \dots - c_1x - c_0$$

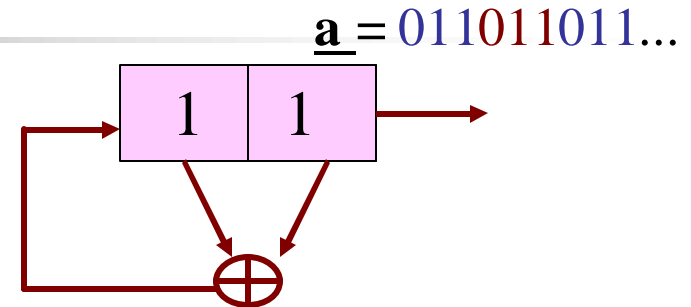
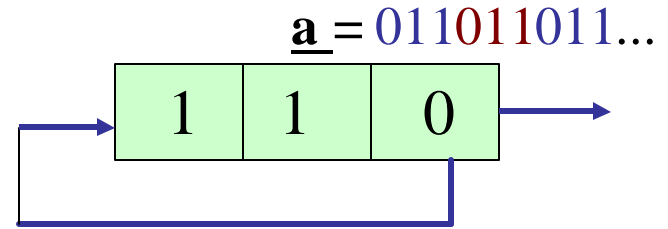
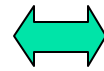
Then $f(x)$ is said to be a characteristic polynomial of \mathbf{a} or the LFSR defined by $g(x_0, x_1, \dots, x_{n-1})$, \mathbf{a} , generated by $f(x)$, and the reciprocal polynomial of $f(x)$, a feedback polynomial.

Example 5 . Let $q = 2$ and

$$f(x) = x^3 + 1.$$

However, \underline{a} can also be generated by an LFSR with a characteristic polynomial:

$$m(x) = x^2 + x + 1$$



A. The minimal polynomials

Definition 1. A monic polynomial with the lowest degree in the set of all characteristic polynomials of \underline{a} is said to be the minimal polynomial (MP) of \underline{a} over F .

Note. The minimal polynomial of the sequence is always a factor of its characteristic polynomials (foundation of parity check attack!).

For Example 5, since $x^3 + 1 = (x+1)(x^2 + x + 1)$ and $x+1$ does not generate \underline{a} , then $x^2 + x + 1$ is the MP of \underline{a} .

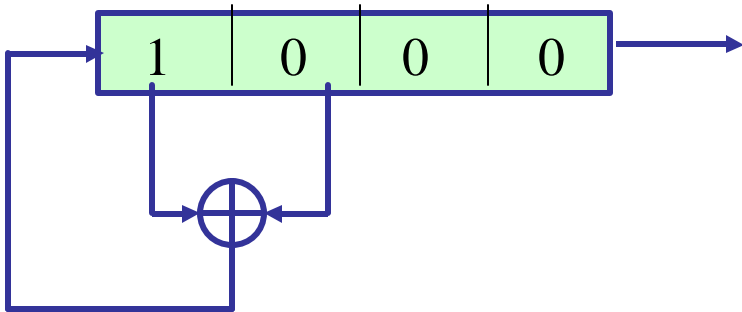
B. Periods

Definition 2. Let $f(x)$ be a polynomial over F . A period of $f(x)$ is defined as the smallest positive integer r such that $f(x) \mid x^r - 1$.

Remark. Periods of output sequences of an LFSR are completely determined by the characteristic polynomial of the LFSR.

Property 1. Let $f(x)$ be the characteristic polynomial of an LFSR, then any output sequence of the LFSR is periodic if and only if $f(0) \neq 0$, i.e., the constant term c_0 of $f(x)$ is nonzero.

Example 6. For the LFSR in Example 4 in Sec 2.1,



$\underline{\mathbf{a}} = 00011011011 \dots$

Corresponding characteristic polynomial $f(x) = x^4 + x^3 + x^2$

Since $f(0) = 0$, then $\underline{\mathbf{a}}$ is not periodic.

C. Irreducible Case

Definition 3. (Shift equivalent) Let $\underline{\mathbf{a}} = a_0, a_1, \dots$ and $\underline{\mathbf{b}} = b_0, b_1, \dots$. If there exists $k \geq 0$, such that

$$b_i = a_{i+k}, i = 0, 1, \dots$$

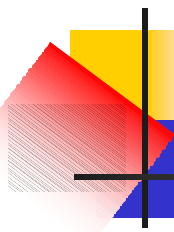
Then we say that $\underline{\mathbf{a}}$ and $\underline{\mathbf{b}}$ are shift equivalent denoted as $\underline{\mathbf{a}} \sim \underline{\mathbf{b}}$. Otherwise, they are shift distinct.

$S = \{\underline{\mathbf{b}} \mid \underline{\mathbf{b}} \sim \underline{\mathbf{a}}\}$, the set consists of all sequences which are shift equivalent with $\underline{\mathbf{a}}$, is called a shift equivalent class of $\underline{\mathbf{a}}$.

If a sequence is periodic with period r , then we may use a vector of dimension r to represent the sequence.

Example 7. For $\underline{\mathbf{a}} = (10001)$, $\underline{\mathbf{b}} = (00011)$, and $\underline{\mathbf{c}} = (11110)$, we have

$\underline{\mathbf{a}} \sim \underline{\mathbf{b}}$, but $\underline{\mathbf{a}}$ and $\underline{\mathbf{c}}$ are shift distinct.



Property 3. Let $f(x)$ be an irreducible polynomial over F of degree n , then the number of shift equivalent classes of non-zero LFSR sequences generated by $f(x)$ is given by

$$(q^n - 1)/\text{per}(f)$$

where $\text{per}(f)$ represents the period of $f(x)$.

Example 8. Let $q = 2$ and $f(x) = x^4 + x^3 + x^2 + x + 1$. Then $f(x)$ is irreducible over $GF(2)$. Determine the cycle structure of the LFSR with $f(x)$ as a characteristic polynomial.

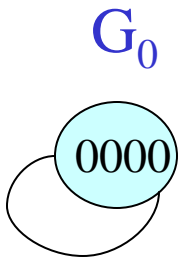
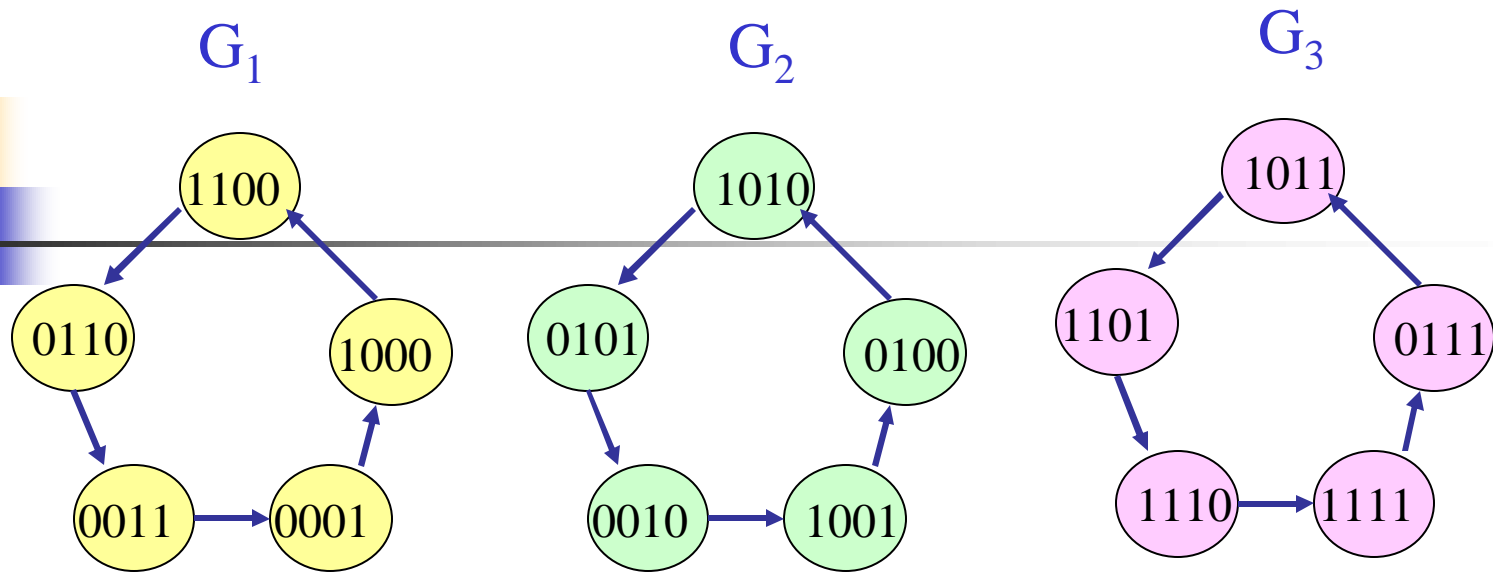
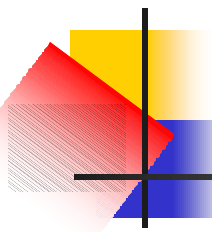
— Note that $x^5 + 1 = (x + 1)(x^4 + x^3 + x^2 + x + 1)$. Then $f(x) \mid x^5 + 1 \Rightarrow \text{per}(f) = 5$.

Thus the set consisting of all LFSR sequences generated by $f(x)$ has

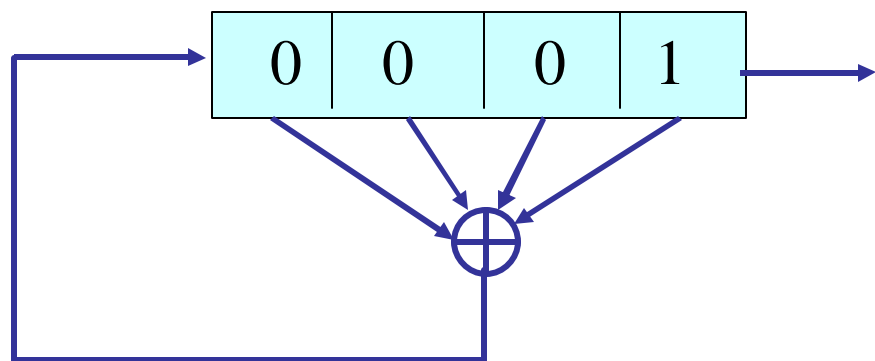
$$15/5 = 3$$

shift equivalent classes for nonzero sequences, which are listed below.

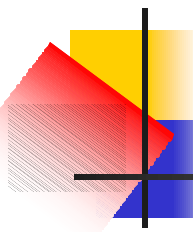
G_0	G_1	G_2	G_3
00000	00011	10010	11110
	00110	00101	11101
	01100	01010	11011
	11000	10100	10111
	10001	01001	01111



State diagram



LFSR with $f(x) = x^4 + x^3 + x^2 + x + 1$



Definition 4. A sequence generated by an LFSR over $F = GF(q)$ with period $q^n - 1$ is called a *maximal length sequence* or an *m-sequence* for short.

Let $f(x)$ be irreducible over F with degree n . If the period of $f(x)$ is $q^n - 1$, then $f(x)$ is said to be *primitive*.

Thus, if $f(x)$ is primitive, then any non-zero sequence generated by $f(x)$ is an *m-sequence*. So, to generate an *m-sequence* is to find a primitive polynomial!

Property 4. If $f(x)$ is primitive, then any nonzero LFSR sequence generated by $f(x)$ has period $q^n - 1$ and all of them are shift equivalent, *i.e.*, the state diagram has two cycles where one contains zero state and the other contains all non states.



Example 9. Let $q = 2$.

- $n = 3$, for the LFSR in Example 2, we have

$$f(x) = x^3 + x + 1$$

then $f(x)$ is primitive, i.e., the period of $f(x)$ is equal to 7.

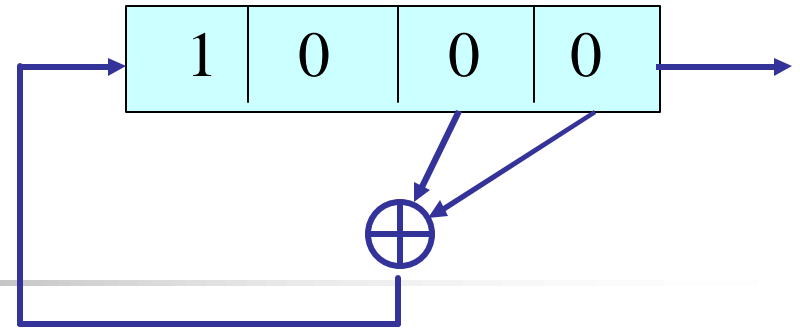
In the following, we use L to represent the shift operator. Then

$$\begin{aligned}\underline{\mathbf{a}} &= 1001011 \\ L\underline{\mathbf{a}} &= 0010111 \\ L^2\underline{\mathbf{a}} &= 0101110 \\ L^3\underline{\mathbf{a}} &= 1011100 \\ L^4\underline{\mathbf{a}} &= 0111001 \\ L^5\underline{\mathbf{a}} &= 1110010 \\ L^6\underline{\mathbf{a}} &= 1100101\end{aligned}$$

All nonzero states of the LFSR are in one cycle.

- $n = 4$, $f(x) = x^4 + x + 1$, primitive over $GF(2)$.

00010 01101 01111, period 15.

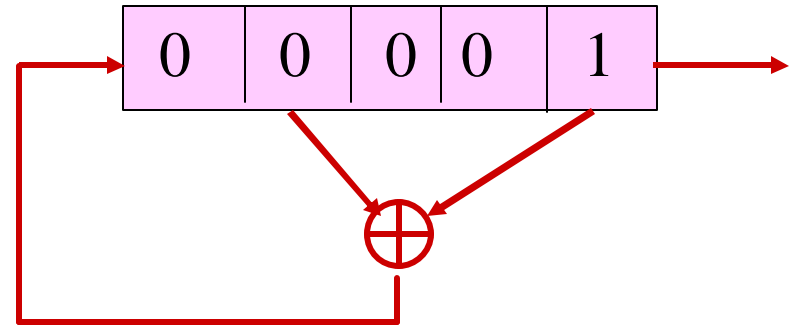


- $n = 5$, $f(x) = x^5 + x^3 + 1$, primitive over $GF(2)$.

1 0 0 0 0 1 0 1 0 1

1 1 0 1 1 0 0 0 1 1

1 1 1 0 0 1 1 0 1 0 0, period 31.



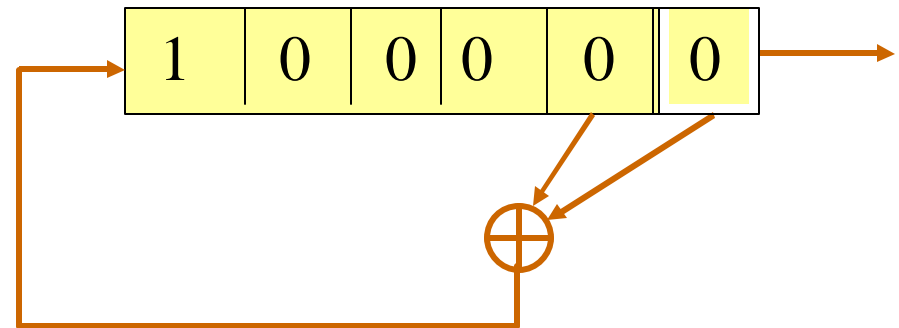
- $f(x) = x^6 + x + 1$, primitive over $GF(2)$.

0 0 0 0 0 1 0 0 0 0 1 1 0 0 0 1 0 1 0 0 1

1 1 1 0 1 0 0 0 1 1 1 0 0 1 0 0 1 0 1 1 0

1 1 1 0 1 1 0 0 1 1 0 1 0 1 0 1 1 1 1 1 1

period 63.





D. Connection with Finite Fields

Trace Representation: i.e., to represent a term of a sequence whose terms taken from $\text{GF}(q)$ by the trace function from $\text{GF}(q^n)$ to $\text{GF}(q)$.

This is a modern approach to design pseudo-random sequences with good correlation.

2. Randomness Measurements

A. Definitions of some basic concepts

Let $\underline{\mathbf{a}} = (a_0, a_1, \dots, a_{N-1})$ be a binary sequence of period N .

Run: For a binary sequence $\underline{\mathbf{a}}$ with period N , k consecutive zeros (ones) is called a run of zeros (or ones) of length k .

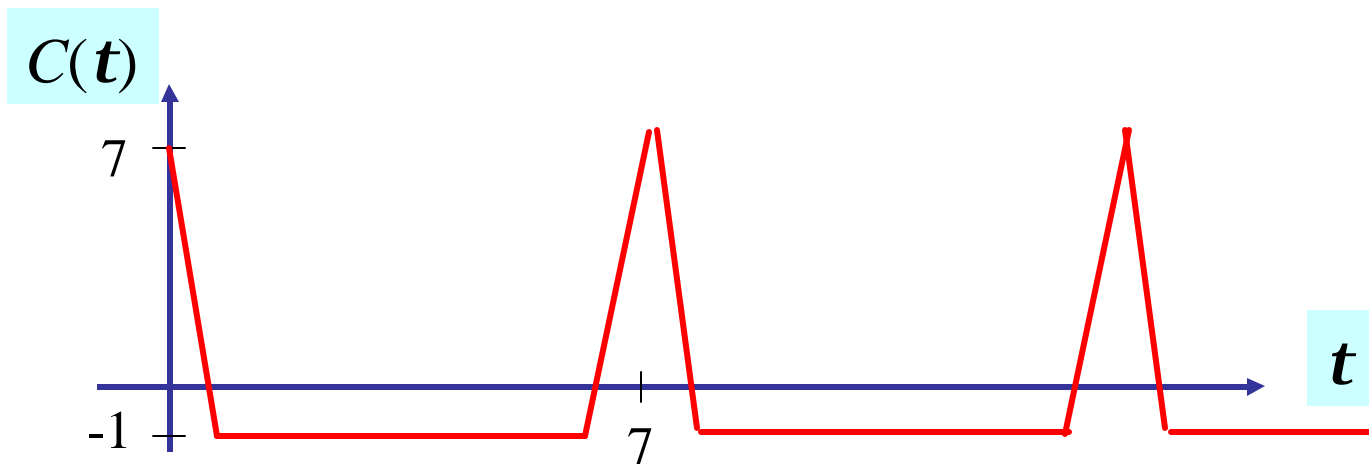
Autocorrelation: The auto correlation function $C(\mathbf{t})$ of $\underline{\mathbf{a}}$ is defined as

$$C(\mathbf{t}) = \sum_{i=0}^{N-1} (-1)^{a_i + a_{i+t}}, \mathbf{t} = 0, 1, \dots.$$

which measures the difference between agreements and disagreements between the sequence and its shifted version.

Example 10. Let $\underline{a} = (1001011)$ with period 7. Then

$$C(t) = \begin{cases} 7 & \text{for } t \equiv 0 \pmod{7} \\ -1 & \text{for } t \not\equiv 0 \pmod{7} \end{cases}$$



Cross correlation between two sequences:

Let $\underline{b} = (b_0, b_1, \dots, b_{N-1})$ be another binary sequence of period N . The cross correlation of \underline{a} and \underline{b} is defined by

$$C_{\underline{a}, \underline{b}}(\tau) = \sum_{i=0}^{N-1} (-1)^{a_i + b_{i+\tau}}$$

Example 11. With \underline{a} in Example 10, let $\underline{b} = (1110100)$. Then

$$C_{\underline{a}, \underline{b}}(0) = 1 \times (-1)^0 + 6 \times (-1) = -5$$

$$C_{\underline{a}, \underline{b}}(2) = 3$$

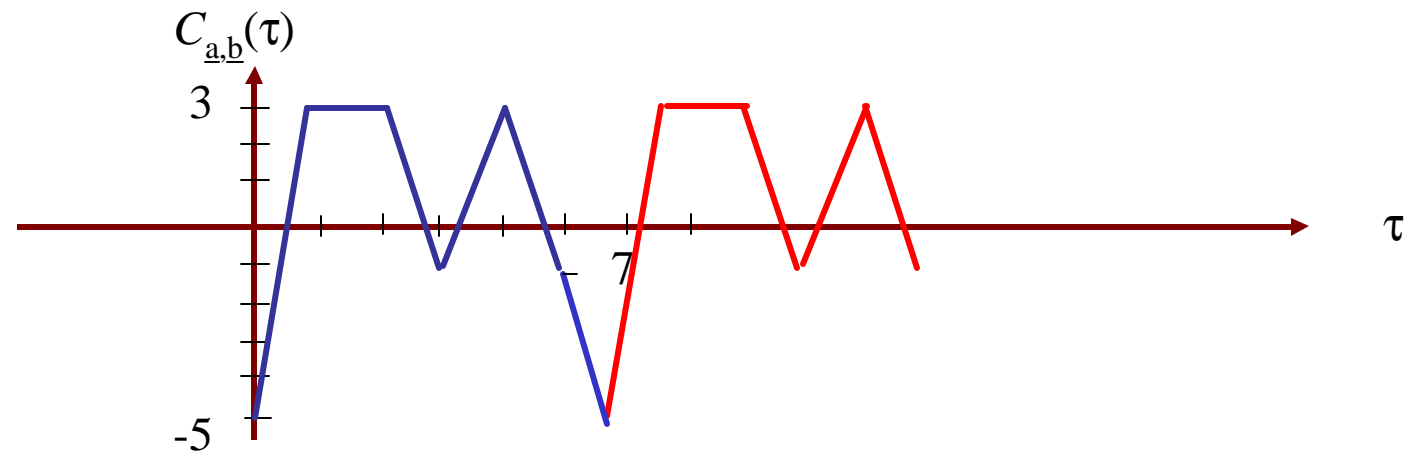
$$C_{\underline{a}, \underline{b}}(4) = 3$$

$$C_{\underline{a}, \underline{b}}(6) = -1$$

$$C_{\underline{a}, \underline{b}}(1) = 5 \times (-1)^0 + 2 \times (-1) = 3$$

$$C_{\underline{a}, \underline{b}}(3) = -1$$

$$C_{\underline{a}, \underline{b}}(5) = -1$$



B. Golomb's Three Randomness Postulates

R1. Balance property: the number of 0's is nearly equal to the number of 1's. Precisely,

$$\left| \sum_{i=0}^{N-1} (-1)^{a_i} \right| \leq 1$$

R2. Run property: In each period, half runs have length 1

1/4 runs have length 2

1/8 runs have length 3

...

Moreover, for each these lengths, there are equally many runs of 0's and that of 1's.

R3. 2-level autocorrelation:

$$C(t) = \begin{cases} N & \text{for } t \equiv 0 \pmod{N} \\ -1 & \text{for } t \not\equiv 0 \pmod{N} \end{cases}$$

C. (Unconditional) Randomness Criteria:

(A) Long period

(B) Statistic Properties:

(1) Balanced property: each element occurs nearly equally many times.

(2) Run property R2.

(3) k -tuple distribution: for $k = \log N$, each k -tuple $d_0, d_1, \dots, d_{k-1}, d_i \in GF(q)$, occurs nearly equally many times in one period.

(C) Correlation

(1) 2-level auto correlation

(2) Low cross correlation: $0 \leq |C_{\mathbf{a}, \mathbf{b}}(\mathbf{t})| \leq \mathbf{s} \sqrt{N}$, where $\mathbf{s} > 0$ is a constant.

(D) Linear span : The degree of the minimal polynomial of $\underline{\mathbf{a}}$ is said to be a linear span of $\underline{\mathbf{a}}$. Thus the linear span of $\underline{\mathbf{a}}$ is the shortest length of LFSR that generates $\underline{\mathbf{a}}$, denoted as $LS(\underline{\mathbf{a}})$. This can be computed by the Berlekamp-Massey algorithm (see Lecture Notes). Large Linear span:

$$Nt / 2 \leq LS(\underline{\mathbf{a}}) \leq N \quad \text{where } 0 < t < 1, \text{ constant.} \quad (\text{B})$$

Let $\mathbf{r} = LS(\underline{\mathbf{a}})/N$ (normalized linear span), then (B) states that $t/2 < \mathbf{r} < 1$.

D. Profiles of Binary M-sequences of Degree n

Period	$2^n - 1$
Balance	2^{n-1} 1's and $2^{n-1} - 1$ 0's
Run property	<p>(1) For $1 \leq k \leq n-2$ runs of 0's (1's) of length k occurs 2^{n-2-k} times</p> <p>(2) Zero run of length $n-1$ occurs once; no runs of 1's of length $n-1$</p> <p>(3) Runs of 1's of length n occurs once</p>
Span n property or ideal n -tuple distribution	Each nonzero n -tuple occurs once
Autocorrelation	2-level
Linear span and r	n , the normalized linear span $r = \frac{n}{2^n - 1}$

As good as it could be!

Only need to know $2n$ bits to reconstruct the entire sequence!

References

- <http://calliope.uwaterloo.ca/~ggong>
- S.W. Golomb and G. Gong, *Design of Signals with Special Properties*, will be delivery to Cambridge University Press soon. The following chapters can be downloaded from the course website of Sequence Design and Cryptography (E&CE 710), Winter 2003, at the above link.
 - Chapter 2. Finite Fields
 - Chapter 3. Feedback Shift Registers Sequences

Other materials related to design of pseudo-random sequence (number) generators can also be downloaded from the course website of E&CE 710, Spring 2002, at the above link.