3.4 The Distribution of Linear Complexity

The distribution of the linear complexities of bit sequences of a fixed length may be exactly determined.

A given sequence $u = (u_0, \ldots, u_{N-1}) \in \mathbb{F}_2^N$ has two possible extensions $\tilde{u} = (u_0, \ldots, u_N) \in \mathbb{F}_2^{N+1}$ by 1 bit. The relation between $\lambda(\tilde{u})$ and $\lambda(u)$ is given by the Massey recursion: Let

$$\delta = \begin{cases} 0 & \text{if the prediction is correct,} \\ 1 & \text{otherwise.} \end{cases}$$

Here "prediction" refers to the next outpit bit from the LFSR we constructed for u. Then

$$\lambda(\tilde{u}) = \begin{cases} \lambda(u) & \text{if } \delta = 0, \\ \lambda(u) & \text{if } \delta = 1 \text{ and } \lambda(u) > \frac{N}{2}, \\ N + 1 - \lambda(u) & \text{if } \delta = 1 \text{ and } \lambda(u) \leq \frac{N}{2}. \end{cases}$$

In the middle case we need a new LFSR, but of the same length.

From these relations we derive a formula for the number $\mu_N(l)$ of all sequences of length N that have a given linear complexity l. To this end let

$$\begin{aligned} M_N(l) &:= \{u \in \mathbb{F}_2^N \mid \lambda(u) = l\} \quad \text{for } N \geq 1 \text{ and } l \in \mathbb{N}, \\ \mu_N(l) &:= \#M_N(l). \end{aligned}$$

The following three statements are immediately clear:

- $0 \leq \mu_N(l) \leq 2^N$,
- $\mu_N(l) = 0$ for $l > N$,
- $\sum_{l=0}^{N} \mu_N(l) = 2^N$.

From these we find explicit rules for the recursion from $\mu_{N+1}(l)$ to $\mu_N(l)$:

Case 1, $0 \leq l \leq \frac{N}{2}$. Every $u \in \mathbb{F}_2^N$ may be continued in two different ways: $u_N = 0$ or 1. Exactly one of them matches the prediction and leads to $\tilde{u} \in M_{N+1}(l)$. The other one leads to $\tilde{u} \in M_{N+1}(N+1-l)$. Since there are no other contributions to $M_{N+1}(l)$ we conclude $\mu_{N+1}(l) = \mu_N(l)$.

Case 2, $l = \frac{N+1}{2}$ (may occur only for odd N). The correctly predicted u_N leads to $\tilde{u} \in M_{N+1}(l)$, however the same is true for the mistakenly predicted one because of the Massey recursion. Hence $\mu_{N+1}(l) = 2 \cdot \mu_N(l)$.

Case 3, $l \geq \frac{N}{2} + 1$. Both possible continuations lead to $\tilde{u} \in M_{N+1}(l)$. Additionally we have one element from each of of the wrong predictions of all $u \in M_{N+1-l}(l)$ from case 1. Hence $\mu_{N+1}(l) = 2 \cdot \mu_N(l) + \mu_{N+1-l}(l)$.

The following lemma summarizes these considerations:

Lemma 14 *The frequency $\mu_N(l)$ of bit sequences of length N and linear complexity l complies with the recursion*

$$\mu_{N+1}(l) = \begin{cases} \mu_N(l) & \text{if } 0 \le l \le \frac{N}{2}, \\ 2 \cdot \mu_N(l) & \text{if } l = \frac{N+1}{2}, \\ 2 \cdot \mu_N(l) + \mu_{N+1-l}(l) & \text{if } l \ge \frac{N}{2} + 1. \end{cases}$$

From this recursion we get an explicit formula:

Proposition 11 [Rueppel] *The frequency $\mu_N(l)$ of bit sequences of length N and linear complexity l is given by*

$$\mu_N(l) = \begin{cases} 1 & \text{if } l = 0, \\ 2^{2l-1} & \text{if } 1 \le l \le \frac{N}{2}, \\ 2^{2(N-l)} & \text{if } \frac{N+1}{2} \le l \le N, \\ 0 & \text{if } l > N. \end{cases}$$

Proof. For $n = 1$ we have $M_1(0) = \{(0)\}$, $M_1(1) = \{(1)\}$, hence $\mu_1(0) = \mu_1(1) = 1$.

Now we proceed by induction from N to $N+1$. The case $l = 0$ is trivial since $M_{N+1}(0) = \{(0, \ldots, 0)\}$, $\mu_{N+1}(0) = 1$. As before we distinguish three cases:

Case 1, $1 \le l \le \frac{N}{2}$. A forteriori $1 \le l \le \frac{N+1}{2}$, and

$$\mu_{N+1}(l) = \mu_N(l) = 2^{2l-1}.$$

Case 2, $l = \frac{N+1}{2}$ (N odd). Here $\mu_N(l) = 2^{2(N-l)}$, and the exponent is $2N - 2l = 2N - N - 1 = N - 1 = 2l - 2$, hence

$$\mu_{N+1}(l) = 2 \cdot 2^{2(N-l)} = 2^{2l-2+1} = 2^{2l-1}.$$

Case 3, $l \ge \frac{N}{2} + 1$. Again $\mu_N(l) = 2^{2(N-l)}$. For $l' = N + 1 - l$ we have $l' \le N + 1 - \frac{N}{2} - 1 = \frac{N}{2}$, hence $\mu_N(l') = 2^{2l'-1}$, and

$$\begin{aligned} \mu_{N+1}(l) &= 2\mu_N(l) + \mu_N(l') = 2^{2N-2l+1} + 2^{2N-2l+1} \\ &= 2^{2N-2l+2} = 2^{2(N+1-l)}. \end{aligned}$$

This completes the proof. ◇

Table 3.1 gives an impression of the distribution.

	1	2	3	4	5	6	7	8	9	10	$N \rightarrow$
0	1	1	1	1	1	1	1	1	1	1	
1	1	2	2	2	2	2	2	2	2	2	
2		1	4	8	8	8	8	8	8	8	
3			1	4	16	32	32	32	32	32	
4				1	4	16	64	128	128	128	
5					1	4	16	64	256	512	
6						1	4	16	64	256	
7							1	4	16	64	
8								1	4	16	
9									1	4	
10										1	
l											
$\downarrow$											

Table 3.1: The distribution of linear complexity

Observations

- Row l is constant from $N = 2l$ on (red numbers), the diagonals, from $N = 2l - 1$ on (blue numbers).
- Each column N, from row $l = 1$ to row $l = N$, contains the powers 2^k, $k = 0, \ldots, N-1$, each one exactly once—first the odd powers in ascending order (red), followed by the even powers (blue) in descending order.
- For every length N there is exactly one sequence of linear complexity 0 and N each: From Section 3.1 we know that these are the sequences $(0, \ldots, 0, 0)$ and $(0, \ldots, 0, 1)$.

Figure 3.5 shows the histogram of this distribution for $N = 10$, Figure 3.6, for $N = 100$. The second histogram looks strikingly small. We'll clarify this phenomen in the following Section 3.5.

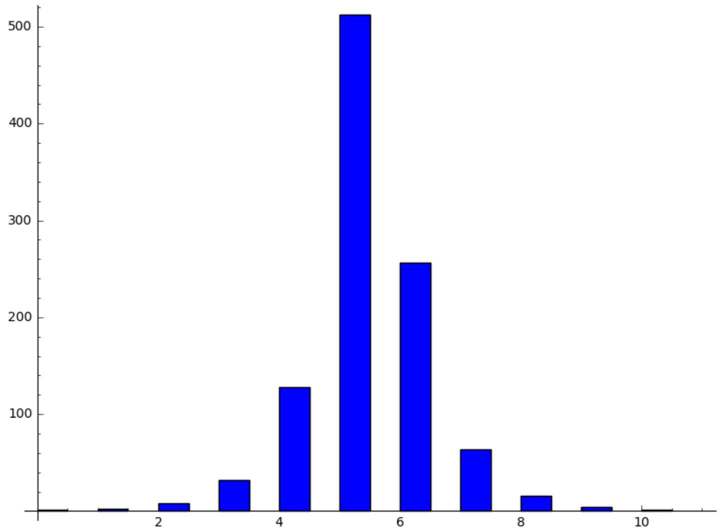

Figure 3.5: The distribution of linear complexity for bitsequences of length $N = 10$

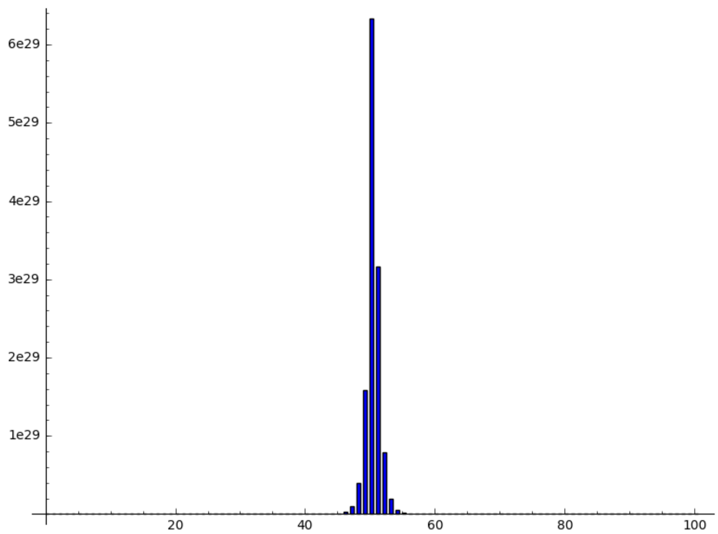

Figure 3.6: The distribution of linear complexity for bitsequences of length $N = 100$