6.4.2 Increasing the Expansion Factor

We now show that the expansion factor of a pseudorandom generator can be increased by any polynomial amount. This means that the previous construction (with expansion factor \( \ell(n) = n + 1 \)) suffices for constructing a pseudorandom generator with arbitrary polynomial expansion factor.

**Theorem 6.23** If there exists a pseudorandom generator \( \hat{G} \) with expansion factor \( \hat{\ell}(n) = n + 1 \), then for any polynomial \( p(n) > n \), there exists a pseudorandom generator \( G \) with expansion factor \( \ell(n) = p(n) \).

**Proof** The idea behind the construction of \( G \) from \( \hat{G} \) is as follows. Given an initial seed \( s \) of length \( n \), the generator \( \hat{G} \) can be used to obtain \( n + 1 \) pseudorandom bits. One of the \( n + 1 \) bits may be output, and the remaining \( n \) bits can be used once again as a seed for \( \hat{G} \). The reason that these \( n \) bits can be used as a seed is because they are pseudorandom, and therefore essentially
as good as a truly random seed. This procedure can be iteratively applied to output as many bits as desired; see Figure 6.1.

We now formally describe the construction of $G$. On input $s \in \{0, 1\}^n$:

1. Let $p'(n) = p(n) - n$. Note that this is the amount by which $G$ is supposed to increase the length of its input.

2. Set $s_0 := s$. For $i = 1, \ldots, p'(n)$ do:
   
   (a) Let $s_{i-1}'$ denote the first $n$ bits of $s_{i-1}$, and let $\sigma_{i-1}$ denote the remaining $i - 1$ bits. (When $i = 1$, $\sigma_0$ is the empty string.)
   
   (b) Set $s_i := (\hat{G}(s_{i-1}'), \sigma_{i-1})$.

3. Output $s_{p'(n)}$.

Before proceeding, note that when $i = 1$, $s_0'$ is the original seed and in step 2b we have $s_1 = \hat{G}(s_0')$. Then, when $i = 2$, the string $s_1$ of length $n + 1$ is split into a prefix of length $n$, denoted $s_1'$, and a suffix of length 1, denoted $\sigma_1$. The string $s_1'$ is used as the seed to $\hat{G}$ again and the resulting string $s_2$ is of length $n + 2$ (namely, it is $(\hat{G}(s_1'), \sigma_1)$). Observe that in the next iteration, the last two bits of $s_2$ become $\sigma_2$ (where the first bit of $\sigma_2$ is the last bit of $\hat{G}(s_1')$ and the second bit of $\sigma_2$ is $\sigma_1$). Thus, in each iteration a single extra bit is generated, and this is incorporated into the “$\sigma$ part”. For this reason, the $\sigma_i$ values grow by one in length in each iteration, as demonstrated in Figure 6.1.

We prove that $G(s)$ is a pseudorandom string of length $p(n)$. We begin by proving this for the simple case of $p(n) = n + 2$.

Define three sequences of distributions $\{H^0_n\}, \{H^1_n\}, \{H^2_n\}$, where each of $H^1_n$, $H^2_n$, and $H^3_n$ is a distribution on strings of length $n + 2$. In distribution

\[\text{FIGURE 6.1: Increasing the expansion of a pseudorandom generator.}\]
That, on input \( w \), the string \( s_0 \leftarrow \{0,1\}^n \) is chosen uniformly at random and the output is \( G(s_0) \). In distribution \( H_0^n \), the string \( s_1 \leftarrow \{0,1\}^{n+1} \) is chosen uniformly at random and then \( G \) is run as above but starting from iteration \( i = 2 \). That is, parse \( s_1 \) as \( (s'_1, \sigma_1) \) with \( |s'_1| = n \), and then output \((G(s'_1), \sigma_1)\). In distribution \( H_1^n \), the string \( s_2 \leftarrow \{0,1\}^{n+2} \) is chosen uniformly at random and output. We denote by \( s_2 \leftarrow H^n_1 \) the choice of the \( (n + 2) \)-bit string \( s_2 \) according to distribution \( H^n_1 \).

We first claim that for any probabilistic polynomial-time distinguisher \( D \) there exists a negligible function \( \text{negl} \) such that

\[
\left| \Pr_{s_2 \leftarrow H^n_0} [D(s_2) = 1] - \Pr_{s_2 \leftarrow H^n_1} [D(s_2) = 1] \right| \leq \text{negl}(n). \tag{6.4}
\]

To see this, fix some \( D \) and consider the polynomial-time distinguisher \( D' \) that, on input \( w \in \{0,1\}^{n+1} \), sets \( s_1 := w \) and then runs \( G \) as above but starting from iteration \( i = 2 \). This yields a string \( s_2 \), and then \( D' \) outputs \( D(s_2) \). The following observations are immediate from the syntactic definitions of \( H^n_0 \) and \( H^n_1 \):

1. If \( w \) is chosen uniformly at random, the distribution on \( s_2 \) generated by \( D' \) is exactly that of distribution \( H^n_1 \). Thus,

\[
\Pr_{w \leftarrow \{0,1\}^{n+1}} [D'(w) = 1] = \Pr_{s_2 \leftarrow H^n_1} [D(s_2) = 1].
\]

2. If \( w = \hat{G}(s) \) for \( s \leftarrow \{0,1\}^n \) chosen uniformly at random, the distribution on \( s_2 \) generated by \( D' \) is exactly that of distribution \( H^n_0 \). I.e.,

\[
\Pr_{s \leftarrow \{0,1\}^n} [D'((\hat{G}(s)) = 1] = \Pr_{s_2 \leftarrow H^n_0} [D(s_2) = 1].
\]

Pseudorandomness of \( \hat{G} \) implies that there exists a negligible function \( \text{negl} \) such that

\[
\left| \Pr_{s \leftarrow \{0,1\}^n} [D'(\hat{G}(s)) = 1] - \Pr_{w \leftarrow \{0,1\}^{n+1}} [D'(w) = 1] \right| \leq \text{negl}(n).
\]

Equation (6.4) follows.

We next claim that for any probabilistic polynomial-time distinguisher \( D \) there exists a negligible function \( \text{negl} \) such that

\[
\left| \Pr_{s_2 \leftarrow H^n_1} [D(s_2) = 1] - \Pr_{s_2 \leftarrow H^n_2} [D(s_2) = 1] \right| \leq \text{negl}(n). \tag{6.5}
\]

The proof is very similar. Consider the polynomial-time distinguisher \( D' \) that, on input \( w \in \{0,1\}^{n+1} \), chooses \( \sigma_1 \leftarrow \{0,1\} \) uniformly at random, sets \( s_2 := (w, \sigma_1) \), and outputs \( D(s_2) \). Notice that if \( w \) is chosen uniformly at
random then \( s_2 \) is uniformly distributed and so is distributed exactly according to \( H_n^2 \). Thus,
\[
\Pr_{w \sim \{0,1\}^{n+1}}[D'(w) = 1] = \Pr_{s_2 \sim H_n^2}[D(s_2) = 1].
\]

On the other hand, if \( w = \hat{G}(s) \) for \( s \sim \{0,1\}^n \) chosen uniformly at random then \( s_2 \) is distributed exactly according to \( H_n^2 \) and so
\[
\Pr_{s \sim \{0,1\}^n}[D'\!(\hat{G}(s)) = 1] = \Pr_{s_2 \sim H_n^2}[D(s_2) = 1].
\]

As before, pseudorandomness of \( \hat{G} \) implies Equation (6.5).

Fix some probabilistic polynomial-time distinguisher \( D \). We have
\[
\Pr_{s \sim \{0,1\}^n}[D(G(s)) = 1] - \Pr_{r \sim \{0,1\}^{n+2}}[D(r) = 1] \quad (6.6)
\]
\[
= \Pr_{s_2 \sim H_n^2}[D(s_2) = 1] - \Pr_{s_2 \sim H_n^2}[D(s_2) = 1] \quad (6.6)
\]
\[
\leq \Pr_{s_2 \sim H_n^2}[D(s_2) = 1] - \Pr_{s_2 \sim H_n^2}[D(s_2) = 1]
\]
\[
+ \Pr_{s_2 \sim H_n^2}[D(s_2) = 1] - \Pr_{s_2 \sim H_n^2}[D(s_2) = 1].
\]

Using Equations (6.4) and (6.5), we conclude that Equation (6.6) is negligible.

**The full proof.** We now give a proof for arbitrary \( p(n) \). The main difference here is a technical one: in the case of \( p(n) = n + 2 \) we could define and explicitly work with three sequences of distributions \( \{H_n^0\}, \{H_n^1\}, \) and \( \{H_n^2\} \). Here, in contrast, we will (in some sense) need to deal with infinitely-many sequences of distributions. Instead of dealing with these explicitly, we deal with them *implicitly* using a common technique known as a *hybrid argument.* (Actually, even the case of \( p(n) = 2 \) utilized a simple hybrid argument.)

Let \( p'(n) = p(n) - n \). For any \( n \) and \( 0 \leq j \leq p'(n) \), let \( H_n^j \) be the distribution on strings of length \( p(n) \) defined as follows: choose \( s_j \sim \{0,1\}^{n+j} \) uniformly at random and then run \( G \) starting from iteration \( i = j + 1 \) and output \( s_{p'(n)} \). (When \( j = p'(n) \) this means we simply choose \( s_{p'(n)} \sim \{0,1\}^{p(n)} \) uniformly at random and output it.) The crucial observation here is that \( H_n^0 \) corresponds to outputting \( G(s) \) for \( s \sim \{0,1\}^n \) chosen uniformly at random, while \( H_n^{p'(n)} \) corresponds to outputting a \( p(n) \)-bit string chosen uniformly at random. Fixing any polynomial-time distinguisher \( D \), this means that
\[
\varepsilon(n) \overset{\text{def}}{=} \Pr_{s \sim \{0,1\}^n}[D(G(s)) = 1] - \Pr_{r \sim \{0,1\}^{p(n)}}[D(r) = 1] \quad (6.7)
\]
\[
= \Pr_{s' \sim \{0,1\}^{p(n)}; H_n^0}[D(s_{p'(n)}) = 1] - \Pr_{s' \sim \{0,1\}^{p(n)}; H_n^{p'(n)}}[D(s_{p'(n)}) = 1].
\]
Our goal is to prove that $\varepsilon$ is negligible, implying that $G$ is a pseudorandom generator.

Fix $D$ as above, and consider the distinguisher $D'$ that does the following when given a string $w \in \{0,1\}^{n+1}$ as input:

1. Choose $j \leftarrow \{1, \ldots, p'(n)\}$ uniformly at random.
2. Choose $\sigma_j \leftarrow \{0,1\}^{J-1}$ uniformly at random.
3. Set $s_j := (w, \sigma_j)$. Then run $G$ starting from iteration $i = j + 1$ and compute $s_{p'(n)}$. Output $D(s_{p'(n)})$.

Analyzing the behavior of $D'$ is more complicated than before, though the underlying ideas are the same. Fix $n$ and say $D'$ chooses $j = J$. If $w \leftarrow \{0,1\}^{n+1}$ was chosen uniformly at random, then $s_j$ is uniformly distributed, and so the distribution on $s_{p'(n)}$ is exactly that of distribution $H_n^J$. That is,

$$
\Pr_{w \leftarrow \{0,1\}^{n+1}} [D'(w) = 1 \mid j = J] = \Pr_{s \leftarrow \{0,1\}^{n+1}} [D(s_{p'(n)}) = 1].
$$

Since each value for $j$ is chosen with equal probability,

$$
\Pr_{w \leftarrow \{0,1\}^{n+1}} [D'(w) = 1] = \frac{1}{p'(n)} \sum_{j=1}^{p'(n)} \Pr_{w \leftarrow \{0,1\}^{n+1}} [D'(w) = 1 \mid j = J]
= \frac{1}{p'(n)} \sum_{j=1}^{p'(n)} \Pr_{s \leftarrow \{0,1\}^{n+1}} [D(s_{p'(n)}) = 1]. \tag{6.8}
$$

On the other hand, say $D'$ chooses $j = J$ and $w = \hat{G}(s)$ for $s \leftarrow \{0,1\}^n$ chosen uniformly at random. Mentally setting $s_{J-1} = (s, \sigma_J)$, we see that $s_{J-1}$ is uniformly distributed and so the distribution on $s_{p'(n)}$ is now exactly that of distribution $H_{n}^{J-1}$. That is,

$$
\Pr_{s \leftarrow \{0,1\}^n} [D'(\hat{G}(s)) = 1 \mid j = J] = \Pr_{s \leftarrow \{0,1\}^{n+1}} [D(s_{p'(n)}) = 1]
$$

and then

$$
\Pr_{s \leftarrow \{0,1\}^n} [D'(\hat{G}(s)) = 1] = \frac{1}{p'(n)} \sum_{j=1}^{p'(n)} \Pr_{s \leftarrow \{0,1\}^n} [D'(\hat{G}(s)) = 1 \mid j = J]
= \frac{1}{p'(n)} \sum_{j=1}^{p'(n)} \Pr_{s \leftarrow \{0,1\}^{n+1}} [D(s_{p'(n)}) = 1]
= \frac{1}{p'(n)} \cdot \sum_{j=0}^{p'(n)-1} \Pr_{s \leftarrow \{0,1\}^{n+1}} [D(s_{p'(n)}) = 1]. \tag{6.9}
$$
(Note that the indices of summation have been shifted in the final step.) We can now analyze how well $D'$ distinguishes outputs of $\hat{G}$ from random:

\[
\left| \Pr_{s \leftarrow \{0,1\}^n} [D'(\hat{G}(s)) = 1] - \Pr_{w \leftarrow \{0,1\}^{n+1}} [D'(w) = 1] \right|
\]

\[
= \frac{1}{p'(n)} \cdot \left| \sum_{j=0}^{p'(n)-1} \Pr_{s' \leftarrow H_n^j} [D(s_{p'(n)}) = 1] - \sum_{j=1}^{p'(n)} \Pr_{s' \leftarrow H_n^j} [D(s_{p'(n)}) = 1] \right|
\]

\[
= \frac{1}{p'(n)} \cdot \left| \Pr_{s' \leftarrow H_n^0} [D(s_{p'(n)}) = 1] - \Pr_{s' \leftarrow H_n^{p'(n)}} [D(s_{p'(n)}) = 1] \right|
\]

\[
= \frac{\varepsilon(n)}{p'(n)}.
\]

relying on Equations (6.8) and (6.9) for the first equality and Equation (6.7) for the final equality. (The second equality is due to the fact that the same terms are included in each sum, except for the first term of the left sum and the last term of the right sum.) Since $\hat{G}$ is a pseudorandom generator, $D'$ runs in polynomial time, and $p'$ is polynomial, we conclude that $\varepsilon$ is negligible.

**The hybrid technique.** The hybrid technique is used in many proofs of security and is a basic tool for proving indistinguishability when a basic primitive is applied multiple times. The technique works by defining a series of hybrid distributions that bridge between two “extreme distributions”, these being the distributions that we wish to prove indistinguishable (in the proof above, these correspond to the output of $G$ and a random string, respectively). To apply the proof technique, three conditions should hold. First, the extreme hybrids should match the original cases of interest (in the proof above, this means that $H_n^0$ was equal to the distribution induced by $G$, while $H_n^{p'(n)}$ was equal to the uniform distribution). Second, it must be possible to translate the capability of distinguishing neighboring hybrids into the capability of breaking some underlying assumption (above, distinguishing $H_n^i$ from $H_n^{i+1}$ was essentially equivalent to distinguishing the output of $\hat{G}$ from random). Finally, the number of hybrids should be polynomial, so the “distinguishing success” is only reduced by a polynomial factor.

**An explicit generator with arbitrary expansion factor.** Let $f$ be a one-way permutation with hard-core predicate $hc$. By combining the construction of Theorem 6.22 (that states that $G(s) = (f(s), hc(s))$ is a pseudorandom generator) with the proof of Theorem 6.23, we obtain that

\[
G(s) = \left( f^{p'(n)}(s), hc \left( f^{p'(n)-1}(s) \right), \ldots, hc(s) \right)
\]

is a pseudorandom generator with expansion factor $p(n) = n + p'(n)$. This generator is known as the Blum-Micali generator.