数字信号处理 I

Digital Signal Processing

第七章 有限字长，采样量化与量化噪声

到目前为止，我们所讨论的数字信号与数字系统还没有涉及到数字量化过程中的精度问题，因此以前各章的内容也可称为无限精度的信号处理。但是，实际上通信中的传输信号都是连续时间信号，所谓数字接收机是首先通过对连续时间信号经过采样信号处理得到的，而连续时间的输入信号在经过采样和A/D变换器转变为有限位长的信号时，会带来量化误差，我们称其为量化噪声；此外，任何一个数字系统的处理和存储都只能是有限字长的数，这当然是有限精度的了，因此也必然会带来一定的误差。因此，归纳起来在数字系统中共有两种因量化而引起的误差效应：

- 采样及A/D变换的量化效应；
- 存储和处理运算过程中的量化和有限字长效应。

近年来，由于采样及A/D变换的量化是在实际研究和应用中最常见的问题，而我们的教材这方面的内容比较单薄，加之我们的授课学时有限。所以，我们将侧重讨论采样量化及量化噪声问题。

本章主要内容

7.1 二进制运算与量化；7.2 采样、A/D变换与量化效应

7.3 量化噪声的统计分析与处理

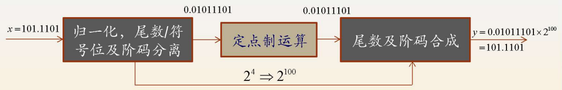
7.1 二进制运算与量化

一、定点与浮点

在整个运算中，小数点在二进制数码中的位置是固定不变的，称为定点制。如M=101.1101，如果这种七位字长的数其小数点始终固定在第三位上，就是定点的。M=101.1101所代表的十进制数为

$$\left(1\times2^{2}+0\times2^{1}+1\times2^{0}\right)+\left(1\times2^{-1}+1\times2^{-2}+0\times2^{-3}+1\times2^{-4}\right)=5.8125$$

通常定点制都把数限制在 ±1，即 $-1<M<1$。这样就把小数点规定在第一位二进制码之前，而把整数位作为“符号位”，代表数的正负号，数的本身只有小数部分，称为“尾数”。定点制的在实际应用中的框图如下：



可以看到：进入定点制之前要归一化和符号位处理，加之要考虑运算中的溢出问题，我们必须引入“阶码”，这里的2^{100}的100就是相应的阶码。

浮点制中为了充分利用尾数的有效位数，总使尾数的第一位保持为1，在浮点制中它称为归一化形式（这不同于定点制前的归一化），例如：$x=0.01011101\times 2^{100}$ 就不是浮点制中的归一化形式；而 $x=0.1011101\times 2^{11}$ 才是浮点制中的归一化形式。所以尾数的范围为 $1/2\le M<1$，这里M是尾数。浮点制是将一个数表述为尾数和指数两部分，如：$x=\pm M\times 2^{c}$，尾数的字长决定了浮点制的运算精度，而阶码c的字长位数决定了浮点制的动态范围。

浮点制的乘法是尾数相乘，阶码相加，尾数相乘过程与定点制相同，因此也要作截尾或舍入的处理。

浮点制的相加需要分三步进行：第一步要对位，使两数的阶码相等；第二步是相加；第三步是使结果归一化并作尾数和阶码处理。例如：$x=0.1010\times 2^{100}$ $y=0.1101\times 2^{010}$ 相加时，首先要将 y 调整到 $y=0.001101\times 2^{100}$，然后再相加

$$\begin{array}{r} 0.1010\quad\times 2^{100} \\ +\ 0.001101\times 2^{100} \\ \hline 0.110101\times 2^{100} \end{array}$$

尾数截尾或舍入处理后得到 $x+y=0.1101\times 2^{100}$。由此可见浮点制的优点是动态范围很大，一般可以不考虑溢出问题。

但在浮点制的运算处理过程中，不论是相乘还是相加，都需要反复考虑归一化和对位及阶码等问题，比定点制复杂的多，运算量也大的多，所以在实际中用的比较少一些。本课程内容主要涉及定点制，侧重讨论采样量化和量化噪声问题。

$$\beta_0.\beta_1\beta_2\cdots\beta_b \qquad (7\text{-}1)$$

二、负数表示法：原码与补码

不论是定点制还是浮点制的尾数都是将整数位用作符号位，其一般的（b+1）位码的形式为：$\beta_0.\beta_1\beta_2\cdots\beta_b$，这里每个 β_i 代表第 i 位二进码，β_i 可以取0或1，β_0 代表 符号位，β_1 至 β_b 代表 b 位字长的尾数值。由于负数表达形式的不同，二进码又可分为原码、补码和反码三种。由于经常使用的就是原码和补码，所以这里我们仅讨论这两种码。

（1） 原码

原码也称为“符号—幅度码”，它的尾数部分代表数的绝对值（即其幅度大小），符号位代表数的正负号，一般 $\beta_0=0$ 代表正数，$\beta_0=1$ 代表负数。例如 $x=0.110$ 表示的是 +0.75，而 $x=1.110$ 则表示的是 -0.75。原码所代表的十进制数值可表示为

$$x=(-1)^{\beta_0}\sum_{i=1}^{b}\beta_i 2^{-i} \qquad (7\text{-}2)$$

原码的优点是乘除运算方便，不论是正负数乘除运算都一样，并以符号位简单地决定结果的正负号。但加减运算则不方便，因为两数相加，先要判断两数符号是否相同。若相同则做加法；若不同则做减法。此时还要判断两数绝对值的大小，以便用大者减小者。

（2） 补码

补码中负数是采用2的补数来表示。也即当 x 为负数时，则用 x 对2的补数 x_c 来代表 x，x_c 的十进位值可按以下公式计算

$$x_c=2-|x| \qquad (7\text{-}3)$$

例如 $x=-0.75$，在原码中表示为1.110，在补码中 $x_c=2-0.75=1.25$，因此补码表示为1.010，这个整数1正好代表了负数。对于一般形式的式（7-1），补码所代表的十进数值可表示为

$$x=-\beta_0+\sum_{i=1}^{b}\beta_i 2^{-i} \qquad (7\text{-}4)$$

例如 补码1.110，按照上式就知道其所表示的数为 $x=-1+0.75=-0.25$

采用补码后，加法运算就方便了，不论数的正负都可直接相加，而且符号位也同样参加运算，如果符号位发生进位，把进位的1丢掉就可以了。

西安交通大学

下面以b=3 为例，列表表示了原码和补码各自所表达的数字。

表7-1 原码和补码的表示法

二进制数	原码值	补码值	二进制数	原码值	补码值
0.111	7/8	7/8	1.000	-0	-1
0.110	6/8	6/8	1.001	-1/8	-7/8
0.101	5/8	5/8	1.010	-2/8	-6/8
0.100	4/8	4/8	1.011	-3/8	-5/8
0.011	3/8	3/8	1.100	-4/8	-4/8
0.010	2/8	2/8	1.101	-5/8	-3/8
0.001	1/8	1/8	1.110	-6/8	-2/8
0.000	0	0	1.111	-7/8	-1/8

由表中可见，每种码均可以组成$\pm 2^3=\pm 8$种数，原码中的0有两个数码表示，因此三位码共能表达$\pm 7/8$以内的15个数值，而在补码中0只有唯一一个表达形式，因此补码的三位码可表达从-1到+7/8之间的16个数值。

西安交通大学

三、量化方式：截尾与舍入

不论是定点制中的乘法还是浮点制的乘法和加法，运算完毕后都会使字长增加。例如原是 b 位字长，运算后增长到 b_1 位字长，因而都需要对尾数作量化处理使 b_1 位字长缩减为 b 位字长。

截尾处理是保留 b 位码，抛掉余下的尾数；而舍入处理则是按接近的值取 b 位码。

这两种处理所产生的误差是不一样的，此外，不同的码制所得结果也不一样。我们来分别加以分析。并且侧重于定点制舍入量化方式。

（1） 定点制的截尾与舍入误差

我们先分析定点制的截尾处理。对于正数，原码和补码的形式都是相同的，即一个b_1位的正数 x 为

$$x=\sum_{i=1}^{b_1}\beta_i 2^{-i} \qquad (7\text{-}6)$$

我们以$[\cdot]$表示量化处理，而以$[\cdot]_T$表示截尾处理，因此

$$[x]_T=\sum_{i=1}^{b}\beta_i 2^{-i} \qquad (7\text{-}7)$$

西安交通大学

以E_T表示截尾误差： $E_T=[x]_T-x=-\sum_{i=b+1}^{b_1}\beta_i 2^{-i}$ $(7-8)$

上式表明截尾误差总是负的，并且在β_i全部为1时，具有最大误差：

$E_T=-\sum_{i=b+1}^{b_1}\beta_i 2^{-i}=-\left(2^{-b}-2^{-b_1}\right)$ ，也即 $-\left(2^{-b}-2^{-b_1}\right)\le E_T\le 0$ 。

一般来说$2^{-b_1}\ll 2^{-b}$ ，并以q 表示2^{-b}，即： $q=2^{-b}$ (7-9)

q 是最小码位所代表的数值，称为“量化宽度”或“量化阶”。因此正数的截尾误差为： $-q<E_T\le 0$ (7-10)

对于负数，原码和补码的表达方式不同，误差也不同。

对于原码负数（$\beta_0=1$）： $x=-\sum_{i=1}^{b_1}\beta_i 2^{-i}$ ，$[x]_T=-\sum_{i=1}^{b}\beta_i 2^{-i}$ ，所以有

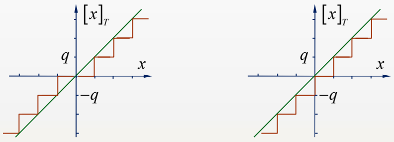
$E_T=[x]_T-x=\sum_{i=b+1}^{b_1}\beta_i 2^{-i}$ ，可见原码负数的误差是正的，即$0\le E_T<q$ 。

例如 b_1 为四位，b 为两位时，负数 x =1.1010 （-0.625），$[x]_T$= 1.10 (-0.5), E_T = $[x]_T$ – x = - 0.5 – (-0.625) = 0.125 > 0 。

对于补码负数（$\beta_0=1$）：$x=-1+\sum_{i=1}^{b_1}\beta_i 2^{-i}$，而 $[x]_T=-1+\sum_{i=1}^{b}\beta_i 2^{-i}$ 。显然这里补码负数的误差与正数一样，仍是负的，且也有 $-q<E_T\leqslant 0$。

例如 x = 1.1001（-0.4375），$[x]_T$= 1.10（-0.5），则

$E_T=[x]_T-x=(-0.5)-(-0.4375)=-0.0625<0$



(a) 原码　　(b) 补码

图7.1 定点制截尾处理的量化特性

总结：原码的截尾误差与数的正负有关，正数时为负，负数时为正，如图7.1（a）所示；而补码的截尾误差都是负值，其特性如图7.1（b）所示。

舍入处理：对于舍入处理，由于是按最接近的数取量化，所以不论是正数、负数，也不论是原码还是补码，其误差总是在±q/2之间。我们以$[\cdot]_R$表示舍入处理。例如

x = 0.1001，$[x]_R$=0.10 舍去0.0001，误差为负-2^{-4}，
x = 0.1011，$[x]_R$=0.11 将0.0011取入为0.01，误差为$+2^{-4}$，
x = 0.1010，则x与0.10及0.11距离相等，因此$[x]_R$既可以取$[x]_R$=0.10，也可以取$[x]_R$=0.11，这一点的选择对误差影响并不大。一般就可以按四舍五入的规则，“逢5进1”，因此取$[x]_R$=0.11。

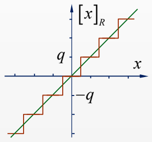


图7.2 定点制舍入处理的量化特性

补码的舍入处理可表示为：$x=-\beta_0+\sum_{i=1}^{b_1}\beta_i 2^{-i}$，$[x]_R=-\beta_0+\sum_{i=1}^{b}\beta_i 2^{-i}+\beta_{b+1}2^{-b}$

最后一项表示逢5进1，其它码也可用类似方式表示。这样我们就得到舍入误差为

$$E_R=[x]_R-x\,,\qquad 而\ -q/2<E_R\leq q/2 \qquad (7-14)$$

(2) 浮点制的舍入误差

在浮点制中截尾或舍入的处理只影响尾数的字长，但是所产生的误差大小却与阶码的值有关。例如x_1和x_2为两个不同阶码的数：

浮点制数	$x_1=0.1001\times 2^{000}\ (=0.5625)$	$x_2=0.1001\times 2^{011}\quad(=4.5)$
量　化	$[x_1]=0.10\times 2^{000}\quad(=0.50)$	$[x_2]=0.10\times 2^{011}\quad(=4.0)$
误　差	$E_1=[x_1]-x_1=-0.0625$	$E_2=[x_2]-x_2=-0.5$

从上面两个数为例我们看到，在同样的尾数舍去的情况下，由于x_2比x_1大8倍，相应的量化误差$|E_2|$也比$|E_1|$大8倍。这说明在浮点制中量化误差是与数字本身的大小有关的，所以用相对误差比用绝对误差更能反映其特点。

相对量化误差定义为 $\varepsilon=\dfrac{[x]-x}{x}$　　(7-15)

据此，绝对误差就可以表示为　$E=[x]-x=\varepsilon x$ 。

分析ε的误差范围。当采用舍入处理时，尾数的误差在±q/2之间，设x的阶码为c，则 $-2^c q/2<[x]_R-x\leq 2^c q/2$，所以有 $-2^c q/2<\varepsilon_R x\leq 2^c q/2$　(7-16)
数x是归一化的浮点数，因此 $2^{c-1}\leq|x|<2^c$　(7-17)
将不等式（7-17）代入不等式（7-16）就可以得到 $-q<\varepsilon_R\leq q$　(7-18)

7.2 采样、A/D变换与量化效应

一、连续、离散与数字信号

信号就是取值随着时间或空间的变化而改变的物理变量。为了能简洁的表述这个概念，除非特殊说明我们总假定该自变量代表时间。如果这个信号的值在一段连续时间上是有效可取的，那么我们称该信号为一个连续时间信号。图7.3给出的就是一个连续时间信号的例子。

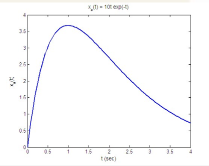


图7.3 连续时间信号 $x_a(t)$

在许多感兴趣的实际应用中，信号仅在一些离散时刻取值有效，这就是我们所说的离散时间信号。也就是说，根据自变量的取值是连续还是离散，可以将信号分为连续时间信号和离散时间信号。

一个产生离散时间信号x(k)的方法，就是通过对连续时间信号进行如下采样：$x(k)=x_a(kT),\quad k=0,1,2,\cdots$ 这里T为采样时间间隔，单位为秒。采样间隔也可以用T的倒数来表示，此时我们称之为采样频率f_s。

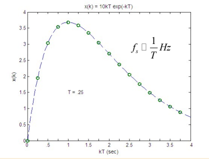


图7.4 采样间隔T=0.25秒的离散时间信号 x(k)

前序课程中大部分的分析都是基于离散时间信号而不是数字信号，也就是说信号值的表示是无限精度的。

这里我们讨论采样量化和数字滤波器问题，必然会涉及到有限精度或有限字长效应。在MATLAB中运行数字滤波器时一般默认为双精度运算，它应对于64Bits精度（十进制的16位字长）。这样高的精度就不会产生明显的有限字长效应。而在字长位数有限的实际系统中，这种字长效应将会变得十分明显，如图7-5所示的数字信号。

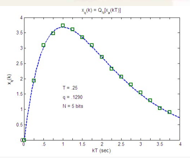


图7.5 数字信号$x_q(k)$

自然界的信号大多是连续（时间）信号，而通常的数字信号都是通过采样和A/D变换量化得到的。对于连续时间信号的采样、A/D量化到数字信号，再由数字处理后通过D/A变换和内插恢复至输出连续时间信号的系统如图7.6所示。

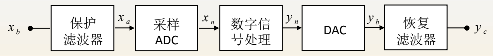


图7.6 具有模拟前置滤波器和后置滤波器的DSP系统

通常这里的保护滤波器和恢复滤波器都必须是模拟滤波器，而且其截至频率应该低于采样频率的二分之一，符合奈奎斯特采样定理。

二、A/D变换的量化效应

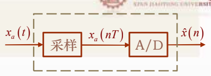


图7.7 采样与A/D变换器

采样和A/D变换从功能上讲一般可以分为两部分，如图7.7所示。通过采样，模拟信号$x_a(t)$转变为采样序列$x_a(nT)$。这里我们侧重于讨论由采样序列$x_a(nT)$到有限位长数字信号x(n)这一转化过程中的量化效应。

定点制补码的截尾与舍入量化效应

为了能清楚地反映这个量化效应，可以把一个实际的A/D变换想象为两个理想的步骤：第一步是无限精度的理想A/D变换，其结果准确的代表采样值，即 $x(n)=Ax_a(nT)$，其中A是比例因子。因为A/D变换总是定点制的，必须使信号不超过A/D变换的动态范围，即 x(n) 必须在（-1,1）之间。x(n)可以用无限长字长来表示，如果采用补码的话，这个表达式为

$$x(n)=Ax_a(nT) \qquad (7-21)$$

$$x(n)=-\beta_0+\sum_{i=1}^{\infty}\beta_i 2^{-i} \qquad (7-22)$$

第二步，则是对 x(n) 量化，使字长固定在b位。这个量化可以采用截尾或者舍入。

对于截尾处理： $$\hat{x}(n)=\left[x(n)\right]_T=-\beta_0+\sum_{i=1}^{b}\beta_i 2^{-i} \qquad (7-23)$$

截尾的量化误差为：$e_T(n)=\hat{x}(n)-x(n)=-\sum_{i=b+1}^{\infty}\beta_i 2^{-i}$

这里如果 $q=2^{-b}$，则 $-q<e_T(n)\le 0 \qquad (7-24)$

这种A/D变换的量化特性如图7.8所示。

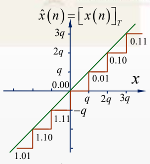


图7.8 补码截尾量化特性

对于舍入处理：

$$\hat{x}(n)=[x(n)]_R=-\beta_0+\sum_{i=1}^{b}\beta_i 2^{-i}+\beta_{b+1}2^{-b} \qquad (7-25)$$

$$e_R(n)=\hat{x}(n)-x(n)$$

考虑到补码舍入误差的正负对称性，如图7.2所示，所以有

$$|e_R(n)|=|\hat{x}(n)-x(n)|=\left|\beta_{b+1}2^{-b}-\sum_{i=b+1}^{\infty}\beta_i 2^{-i}\right|\le\frac{1}{2}\times 2^{-b}$$

这里如果 $q=2^{-b}$，则 $-q/2<e_R(n)\le q/2 \qquad (7-26)$

补码舍入量化特性如图7.9所示，可见补码舍入的量化特性比较符合我们要求。在实际中应用较多。

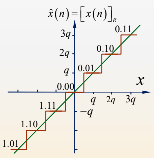


图7.9 补码舍入量化特性

7.3 量化噪声的统计分析与处理

一、量化效应的统计分析

式（7-24）和（7-26）虽然分析了量化误差的范围，但是要完全精确的知道误差究竟多大几乎是不可能的。因为这要看信号的具体情况而定；同时这也没有必要。一般我们只要知道误差的一些平均效应就够了，就可以作为我们设计的依据了。例如由此可以确定A/D变换所需字长、选择A/D芯片、滤波电路结构以及确定实际的采样速率等的依据，所以对于量化误差采用统计分析的方法是合适的。

量化误差e(n)统计模型的一些假定：

（1）e(n)是一个平稳的随机序列；

（2）e(n)是与信号序列x(n)不相关的；

（3）e(n)具有均匀等概的分布；

（4）e(n)序列本身的任意两个值之间也是不相关的，即e(n)是白噪声序列。

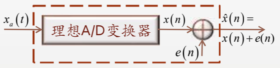


图7.10 A/D变换的统计分析模型

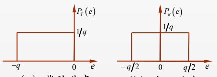
根据这样的假定，量化误差就是一个与信号序列完全不相关的**白噪声序列**，因此也称为量化噪声，它与信号的关系是加性的。这样，一个实际的A/D变换就可以看作为一个理想的A/D变换并在其输出端加入了一个白色噪声序列源e(n)。

应该注意到：这种统计假设并不一定符合实际情况，例如输入$x_a(t)$是直流或者方波这一类规则信号时，显然误差不能认为是线性独立（正交），也不能认为其功率谱是白色的，也就不能使用这一模型。但是对于大多数不规则的自然信号来说，这种假设就非常接近实际。因此作为一种平稳随机信号的概率统计特性分析来说，这些假设是合适的。

作为**白噪声序列**，我们来计算一下e(n)的均值 m_e 和方差 σ_e^2

$$m_e=E[e(n)]=\int_{-\infty}^{\infty}eP_1(e)de \qquad (7-27)$$

$$\sigma_e^2=E\left[(e(n)-m_e)^2\right]=\int_{-\infty}^{\infty}(e-m_e)^2P_1(e)de \qquad (7-28)$$



（a）截尾噪声　（b）舍入噪声

图7.11 量化噪声的概率分布

其中E[.]表示取数学期望，e(n)由于是平稳的，在求数学期望时与n无关，所以可以不标n，$P_1(e)$是误差e值的概率密度。由于e(n)是均匀等概分布，因此对截尾误差及舍入误差，其概率密度分别如图7.11所示。将此概率密度代入以上两式，就可以得到：

截尾量化噪声：$\begin{cases}m_e=-q/2\\ \sigma_e^2=q^2/12\end{cases} \qquad (7-29)$　　舍入量化噪声：$\begin{cases}m_e=0\\ \sigma_e^2=q^2/12\end{cases} \qquad (7-30)$

量化噪声的方差与A/D变换的字长直接相关，字长越长，q越小，量化噪声越小。

西安交通大学 XI'AN JIAOTONG UNIVERSITY

例如：字长b=10时，$q^2=2^{-2b}=2^{-20}$，量化噪声的方差 $\sigma_e^2=7.95\times10^{-8}$，A/D变换器输出信号的最大绝对值不超过1。因此$\sigma_e^2$ 比最大信号值低71dB（$=-10\log_{10}(\sigma_e^2)$）。当字长增加到15位时，$\sigma_e^2=7.76\times10^{-11}$，这时$\sigma_e^2$ 就比最大信号值低101dB。当然字长越长A/D变换器的信噪比越高。但字长过长也没有必要，因为输入信号$x_a(t)$本身有一定的信噪比，A/D变化的量化阶q比$x_a(t)$的噪声电平低的多是没有意义的。

另外，我们看到截尾噪声具有直流分量，将影响信号的频谱结构，因此一般总是更愿意采用舍入量化处理。我们以后也只讨论舍入量化。

二、量化误差的时域（统计）表达

（1）数学期望 $m_e=E\left[e(n)\right]=\int_{-\infty}^{\infty}eP_1(e)de$ 　　$m_e \Rightarrow$ 直流分量，$m_e^2 \Rightarrow$ 直流功率。

（2）均方值 $E\left[e^*(n)e(n)\right]=E\left[\left|e(n)\right|^2\right]=\int_{-\infty}^{\infty}|e|^2P_1(e)de$ 　　“总功率”或“平均功率”

（3）方差 $\sigma_e^2=E\left[\left(e(n)-m_e\right)^*\left(e(n)-m_e\right)\right]=E\left[\left|\left(e(n)-m_e\right)\right|^2\right]=\int_{-\infty}^{\infty}\left|\left(e-m_e\right)\right|^2P_1(e)de$

方差是“交流功率”，　总功率=直流功率+交流功率=平均功率。

西安交通大学 XI'AN JIAOTONG UNIVERSITY

因此有 $E\left[\left|e(n)\right|^2\right]=m_e^2+\sigma_e^2 \quad\Rightarrow\quad \sigma_e^2=E\left[\left|e(n)\right|^2\right]-m_e^2$

这三者（总功率、直流功率和交流功率）都只和一维概率密度有关。而对于平稳随机量化噪声而言，概率密度与时间n是无关的。以下会涉及到二维概率。

（4）自相关函数

$$\phi_{ee}(m)=E\left[e^*(n_1)e(n_2)\right]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e_1^*e_2P_2(e_1,e_2,m)de_1de_2=E\left[e^*(n)e(n+m)\right]$$

（5）自协方差函数

$$\gamma_{ee}(m)=E\left[\left(e(n_1)-m_e\right)^*\left(e(n_2)-m_e\right)\right]=E\left[\left(e(n)-m_e\right)^*\left(e(n+m)-m_e\right)\right]$$

$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(e_1-m_e)^*(e_2-m_e)P_2(e_1,e_2,m)de_1de_2$$

$\therefore\ \gamma_{ee}(m)=\phi_{ee}(m)-m_e^2$ 　这里误差均值 $m_e=0$，所以 $\gamma_{ee}(m)=\phi_{ee}(m)$

量化误差e(n)是一个平稳的随机过程，所以其均值m_e，均方值$E[|e|^2]$和方差σ_e^2均与n无关。而自相关函数和自协方差函数中的$m=n_2-n_1$，是时间差的概念。所以可以看出自相关函数$\phi_{ee}(m)$和自协方差函数$\gamma_{ee}(m)$均为m的函数。实际上，考虑到平稳随机过程的“各态历经”的假设：集合的平均就等于时间的平均。据此，“相关”在实际中就是我们以前讲的“内积”。

西安交通大学 XI'AN JIAOTONG UNIVERSITY

因为 $\gamma_{ee}(m)=\phi_{ee}(m)$，而当 $n_1=n_2=n$ 时，$m=n_2-n_1=n-n=0$，所以有

$$\gamma_{ee}(0)=\phi_{ee}(0)=E\left[e^*(n)e(n)\right]=E\left[\left|e(n)\right|^2\right]=\sigma_e^2$$

量化噪声序列e(n)是一个零均值的白噪声序列，所以其自相关函数应为

$$\phi_{ee}(m)=\sigma_e^2\delta(m)$$

根据Wiener-Khinchin定理：自相关函数与功率谱是一对傅里叶变换对，即

$$\begin{cases}P_{ee}(\omega)=\sum_{m=-\infty}^{\infty}\phi_{ee}(m)e^{-j\omega m}, & \Rightarrow\quad P_{ee}(\omega)=\sigma_e^2\\ \phi_{ee}(m)=\frac{1}{2\pi}\int_{-\pi}^{\pi}P_{ee}(\omega)e^{j\omega m}d\omega, & \Rightarrow\quad \phi_{ee}(m)=\sigma_e^2\delta(m)\end{cases}$$

而我们应该注意到：这里的ω是数字角频率，它与采样前所用的模拟角频率Ω是有关系的 ω=ΩT。据此，我们代入这一关系式，有

$$\phi_{ee}(m)=\frac{T}{2\pi}\int_{-\frac{\pi}{T}}^{\frac{\pi}{T}}P_{ee}(\Omega T)e^{j\Omega Tm}d\Omega=\sigma_e^2\delta(m)$$

这就说明:噪声谱密度不仅与量化台阶大小有关，也与采样频率$f_s=1/T$有关系。

同时我们前边已经看到量化误差的方差与量化台阶大小有关 　$\sigma_e^2=q^2/12$

三、量化噪声通过系统

在讨论量化噪声通过线性系统后的影响时，我们可以近似地将系统看做是完全理想的，即无限精度的线性系统。因此线性相加的输入噪声在系统的输出端仍然是线性相加的，如图7.12所示。

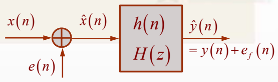


图7.12 量化噪声通过线性系统

这样，系统的输出为

$$\hat{y}(n)=\hat{x}(n)*h(n)=\left[x(n)+e(n)\right]*h(n)$$
$$=x(n)*h(n)+e(n)*h(n)$$

因此，输出噪声可以表示为 $e_f(n)=e(n)*h(n)$

这里e(n)是舍入噪声，所以输出噪声的方差为

$$\sigma_f^2=E\left[\left|e_f(n)\right|^2\right]=E\left\{\left[\sum_{m=0}^{\infty}h(m)e(n-m)\right]^*\left[\sum_{l=0}^{\infty}h(l)e(n-l)\right]\right\}$$
$$=\sum_{m=0}^{\infty}\sum_{l=0}^{\infty}h^*(m)h(l)E\left[e^*(n-m)e(n-l)\right]$$

由于e(n)是白色的，即e(n)序列的各变量之间互不相关，因此有

$$E\left[e^*(n-m)e(n-l)\right]=\sigma_e^2\delta(m-l)$$

$$E\left[e^*(n-m)e(n-l)\right]=\sigma_e^2\delta(m-l)$$

将这一结果代入以上的输出噪声方差式

$$\sigma_f^2=\sum_{m=0}^{\infty}\sum_{l=0}^{\infty}h^*(m)h(l)\cdot\sigma_e^2\delta(m-l)=\sigma_e^2\sum_{m=0}^{\infty}\left|h(m)\right|^2 \qquad (7-31)$$

根据Parserval定理，频域的总能量等于时域的总能量，所以式（7-31）也可以表示为

$$\sigma_f^2=\frac{\sigma_e^2}{2\pi j}\oint_c H(z)H(z^{-1})\frac{dz}{z}=\frac{\sigma_e^2}{2\pi}\int_{-\pi}^{\pi}\left|H(e^{j\omega})\right|^2 d\omega \qquad (7-33)$$

由上一小节的式子 $\phi_{ee}(m)=\frac{T}{2\pi}\int_{-\frac{\pi}{T}}^{\frac{\pi}{T}}P_{ee}(\Omega T)e^{j\Omega Tm}d\Omega=\sigma_e^2\delta(m)$ ，可以得出

$$\sigma_f^2=\frac{\sigma_e^2 T}{2\pi}\int_{-\frac{\pi}{T}}^{\frac{\pi}{T}}\left|H(e^{j\Omega T})\right|^2 d\Omega \qquad (7-34)$$

对于舍入量化噪声通过系统后的数学期望或均值而言:

$$m_f=E\left[e_f(n)\right]=E\left[\sum_{m=0}^{\infty}h(m)e(n-m)\right]=\sum_{m=0}^{\infty}h(m)E\left[e(n-m)\right]=m_e\sum_{m=0}^{\infty}h(m)$$

因为舍入量化噪声的均值为零，所以 $m_f=m_e\sum_{m=0}^{\infty}h(m)=0$ ，**输出的均值也为零**。

四、采样及量化效应原理在实际中的应用

（1）通用的中频采样与A/D变换量化系统

例如：中频的中心频率为f_I=70MHz，中频带宽为10MHz，也就是说中频频带是从65MHz到75MHz，如图7.13所示。所以采样频率定为f_s=160MHz，8位的A/D变换器。整个的通用的中频采样与A/D变换量化系统框图如图7.14所示。

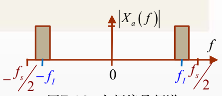


图7.13 中频信号频谱及采样速率示意图

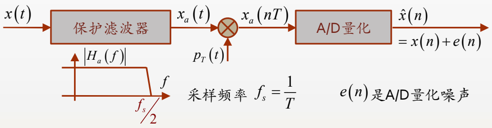


图7.14 通用中频采样与A/D变换量化系统框图

首先我们来分析该系统这一系统的量化噪声e(n)的方差： $\sigma_e^2=q^2/12$

对于8位A/D变换器而言，量化阶q为：$q=2^{-8}$，而 $q^2=2^{-16}$，量化噪声方差为

$$\sigma_e^2=2^{-16}/12=1.273\times10^{-6}$$

8位A/D变换器的量化方差： $\sigma_e^2 = 2^{-16}/12 = 1.273\times10^{-6}$

而如果我们将A/D变换器换成10位A/D变换器，则 $\sigma_e^2 = 2^{-20}/12 = 7.947\times10^{-8}$ 。可见，10位的A/D变换器的量化噪声方差会比8位的小的多了。量化噪声的方差是衡量A/D变换器水平主要指标。

（2）不换A/D芯片，不改变采样速率，改进中频采样A/D量化性能的措施

以上我们已经看到：量化噪声的方差是衡量A/D变换器水平主要指标。要想不换A/D变换芯片，改善A/D量化的性能，就可以从减小量化噪声的方差着手。可以通过对图7.14中频采样与A/D变换量化噪声输出e(n)的滤波处理来减小量化噪声，而这种滤波不应伤及信号x(n)。据此，我们采用如下滤波处理.

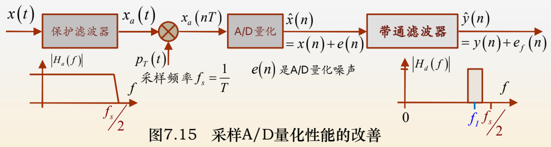


图7.15　采样A/D量化性能的改善

这里我们仍采用8位A/D芯片以160MHz速率采样；为了不伤及信号，带通滤波器的

截至频率为65MHz-75MHz。这样式（7-33）和式（7-34）可以得出带通滤波后量化噪声$e_f(n)$的方差

$$\sigma_f^2 = \frac{\sigma_e^2}{2\pi}\int_{-\pi}^{\pi}\left|H_d\left(e^{j\omega}\right)\right|^2 d\omega, \qquad \text{注意到}\ \omega=\Omega T,\ \Omega=\frac{\omega}{T}=\omega f_s,\ d\omega = Td\Omega$$

$$= \frac{\sigma_e^2 T}{2\pi}\int_{-\frac{\pi}{T}}^{\frac{\pi}{T}}\left|H_d\left(e^{j\Omega T}\right)\right|^2 d\Omega, \qquad \text{考虑到图7.15中带通滤波器的特性，则}$$

$$= \frac{\sigma_e^2}{2\pi f_s}\int_{-\pi f_s}^{\pi f_s}\left|H_d\left(e^{j\Omega T}\right)\right|^2 d\Omega = \frac{\sigma_e^2}{2\pi f_s}\left\{\int_{-\frac{75}{80}\pi f_s}^{-\frac{65}{80}\pi f_s} d\Omega + \int_{\frac{65}{80}\pi f_s}^{\frac{75}{80}\pi f_s} d\Omega\right\} = \frac{\sigma_e^2}{2\pi f_s}\cdot\frac{2\pi f_s}{8}$$

所以有： $\sigma_f^2 = \frac{\sigma_e^2}{8}$

这意味着：由于带通滤波的作用，使得量化噪声的功率减小了7/8，A/D量化器的性能大为改善了。这样的改善等价于我们采用了多少位的A/D变换器呢？下面我们来分析一下：量化噪声方差 $\sigma_e^2 = q^2/12$ ，其中 $q = 2^{-N}$ ，所以现对于8位A/D而言，

$$\frac{2^{-2N}}{\cancel{12}} = \sigma_f^2 = \frac{\sigma_e^2}{8} = \frac{\left(q^2/12\right)}{8} = \frac{\left(2^{-8}\right)^2}{8\times12} = \frac{2^{-16}}{8\times\cancel{12}} \quad\Rightarrow\quad 2^{-2N} = 2^{-19} \quad\Rightarrow\quad N = 9.5$$

N=9.5这意味着我们在没有更换A/D芯片，也没有改变采样速率的条件下，用8位A/D芯片达到了9.5位A/D芯片的性能。这当然是非常有意义的。

（3）改变采样速率，改进中频采样A/D量化性能的措施

目前采样速率在500MHz以下的8位A/D芯片较为通用，市场售价便宜，性价比高。据此，我们可以在（2）中改进的基础上，通过适当地提高采样速率（不超过500MHz），例如：从原来的160MHz提高到320MHz，从而实现A/D量化性能的进一步改善。这样带通滤波的截至频率仍为65MHz-75MHz。量化噪声$e_f(n)$的方差为

$$\sigma_f^2 = \frac{\sigma_e^2}{2\pi f_s}\int_{-\pi f_s}^{\pi f_s}\left|H_d\left(e^{j\Omega T}\right)\right|^2 d\Omega = \frac{\sigma_e^2}{2\pi f_s}\left\{\int_{-\frac{75}{160}\pi f_s}^{-\frac{65}{160}\pi f_s} d\Omega + \int_{\frac{65}{160}\pi f_s}^{\frac{75}{160}\pi f_s} d\Omega\right\} = \frac{\sigma_e^2}{2\pi f_s}\cdot\frac{2\pi f_s}{16}$$

所以有： $\sigma_f^2 = \frac{\sigma_e^2}{16} = 2^{-4}\sigma_e^2$

量化噪声的功率减小到原来的1/16了，A/D量化性能大为改善。也计算等价位数

$$\frac{2^{-2N}}{\cancel{12}} = \sigma_f^2 = \frac{\sigma_e^2}{16} = \frac{\left(q^2/12\right)}{2^4} = \frac{2^{-16}}{2^4\times\cancel{12}} \quad\Rightarrow\quad 2^{-2N} = 2^{-20} \quad\Rightarrow\quad N = 10$$

N=10这意味着：我们通过适当提高采样速率和设置带通滤波器，用8位A/D芯片达到了10位A/D芯片的性能。这时的量化噪声方差已不再是8位的1.273×10^{-6}，而是10位对应的7.947×10^{-8}了，量化噪声的功率减小了两个数量级。

采样、量化及量化噪声小结

（1）A/D量化由于有限位数、有限精度的限制会产生量化噪声，这种舍入量化噪声是一种零均值的白噪声。

（2）量化噪声的方差（功率）=$q^2/12$，其中q是量化阶（量化台阶），$q=2^{-N}$，N是A/D量化的位数。

（3）A/D量化后设置的带通数字滤波器的通带与中频信号的带宽匹配，主要用于滤除带外量化噪声，也可以滤除其他带外噪声。任何模拟滤波器都不可能取代其作用。

（4）由于数字带通滤波器的设置，使得提高采样速率改善采样与A/D量化性能成为可能。