第 8 卷 第 6 期
2019 年 11 月
集 成 技 术
JOURNAL OF INTEGRATION TECHNOLOGY
Vol. 8 No. 6
Nov. 2019

引文格式：
阮伟华, 王海鹏. 应用于 100GE 高速变速箱设计方法及时序分析 [J]. 集成技术, 2019, 8(6): 1-10.
Ruan WH, Wang HP. Design method and timing analysis of the high-speed gearbox for 100GE [J]. Journal of Integration Technology, 2019, 8(6): 1-10.

应用于 100GE 高速变速箱设计方法及时序分析

阮伟华 王海鹏

（三江学院电子信息工程学院 南京 210012）

摘 要 该文介绍了 4 种变速箱设计方法，并对它们进行时序分析。通过对比单元数量（面积）、功耗、速度及稳定性，选择了一种基于轮循存储方式的变速箱应用到 100GE 物理编码子层电路中。该变速箱可以在一段时间范围内开始输出，克服了输入输出时钟相位差的影响，提高了电路的速度和稳定性。经过结构优化及流水线结构设计，该变速箱的时钟速度超过 700 MHz。另外，采用 0.18 μm 互补金属氧化物半导体工艺对物理编码子层电路进行流片，测试结果表明该电路能够以 100 Gb/s 速率稳定工作，进一步证明了变速箱设计的正确性。

关键词 100GE；物理编码子层；半定制；变速箱；轮循
中图分类号 TN 47 **文献标志码** A doi: 10.12146/j.issn.2095-3135.20190423001

Design Method and Timing Analysis of the High-Speed Gearbox for 100GE

RUAN Weihua WANG Haipeng

(*Electronic Information Engineering College*, *Sanjiang University*, *Nanjing* 210012, *China*)

Abstract In this paper, four design methods of gearbox are introduced, and their timing analysis is carried out. By comparing the number of cells (area), power consumption, speed and stability, a gearbox based on round-robin saving mode is selected and applied to 100GE physical coding sublayer circuit. It can take out the output value within a certain range, which overcomes the influence of phase difference between input and output clocks and greatly improves the speed and stability of the circuit. After structure optimization and pipeline design, the gearbox can work stability at the clock frequency of over 700 MHz and meet the design requirement. The physical coding sublayer circuit which including the gearbox has been taped out in 0.18 μm complementary metal oxide semiconductor technology and measured results show that it can work properly at speed of 100 Gb/s.

Keywords 100GE; physical coding sublayer; semicustom; gearbox; round-robin

收稿日期：2019-04-23 **修回日期**：2019-07-21
基金项目：国家自然科学基金项目（61801262）；江苏省高等学校自然科学研究项目（18KJB10039）；三江学院科研资助项目（2018SJKY012）
作者简介：阮伟华（通讯作者），博士，副教授，研究方向为高速数字 IC 研究，E-mail：rwh111@163.com；王海鹏，博士，研究方向为生物芯片。

1 引 言

以太网具有成本低、可靠性高、安装维护简单等优点，是目前普遍采用的一种网络技术。随着互联网技术的不断发展和用户数量的不断增加，用户对数据传输和接入带宽的需求将越来越大。于是在 2010 年制定了 40G/100G 以太网的标准 IEEE802.3ba[1]。其实在此标准制定之前，人们就着手研究 100G 以太网。比如，研究人员[2-4]分别介绍了 100GE 的物理编码子层(Physical Coding Sublayer，PCS)结构、光纤局域网技术及 100 Gb/s 的差分正交相移键控(DQPSK)传输实验。随后出现不少关于 100GE 的研究文献，如介绍用现场可编程门阵列(FPGA)实现 100GE 的物理(PHY)和介质访问控制(MAC)层及采用 FPGA 实现 100GE 网络监测等[5-7]；还有一些文献介绍 100GE 的测试、前向错误校正(FEC)和密集波分复用(DWDM)传输系统，或介绍具有相位调制的 100GE 可行性研究等[8-13]。但这些文献均没有涉及 100GE 物理层的专用集成电路(ASIC)实现，特别是用 0.18 μm CMOS(互补金属氧化物半导体)工艺的实现。

本文对 100GE 物理编码子层(PCS)中的一个关键模块——变速箱模块进行研究和设计，并对 4 种不同的变速箱结构进行了研究和时序分析。其中，变速箱是一种把低速、位宽大的数据转换成高速、位宽小的数据，或者反过来，转换前后数据传送率保持不变的电路。例如，m∶n 变速箱(其中 m、n 分别代表输入输出信号的位宽，且 m 不等于 n)输入输出时钟频率的比值为 n∶m。Ruan 和 Hu[4]介绍的 100GE 发送端 PCS 电路中采用了 66∶8 变速箱，其目的是把输入速度为 78.125 MHz 的 66 比特块转换成输出速度为 644.531 25 MHz 的 8 比特数据块。由于输入输出的时钟频率不一致，变速箱极易受到输入输出时钟相位差的影响，因此设计一个高效并绝对稳定可靠的变速箱是整个 PCS 层电路的关键。本文设计了 4 种不同结构的变速箱，并对它们的功耗面积以及时序进行了详细分析；然后采用其中最优的结构应用于 PCS 电路中，并对 PCS 电路进行芯片设计并流片和测试。测试结果表明，所采用的变速箱高效并且 100% 数据稳定。

2 基于两级移位寄存器的 66∶8 变速箱

在介绍变速箱之前，首先简单了解一下图 1 中 PCS 电路的结构[14]。该电路是根据文献[1,5]设计的一个电气接口采用 4×25 Gb/s 的 100GE 发送端 PCS 电路，其主要包括数据发生器、64B/66B 编码器、扰码器(256 bit)、多通道分发(MLD)和 66∶8 变速箱。其中，数据发生器的加入是为了测试的需要：当电路复位后，数据发生器模拟调和子层(Reconciliation Sublayer) [1]不断地产生 4 路平行的 64 比特的数据字 CGMII_TXD (4×64 b×390.625 MHz＝100 Gb/s)及 8 比特的控制字 CGMII_TXC 这两种伪随机码。首先，64B/66B 编码器把这 4 路 CGMII_TXD 和 CGMII_TXC 编码成 4 路平行的 66 比特块(block)；然后，经扰码器进行扰码(其中 2 比特的同步头不用扰码)，再经多通道分发模块(MLD)把扰码后的 66 比特块轮循(round-robin)分发成 20 路虚拟通道；接着，66∶8 变速箱把这 20 路速度为 78.125 MHz 的 66 比特块转换成 20 路速度为 644.531 25 MHz 的 8 比特数据；最后，这 20 路平行的 8 比特数据被两级复接器(8∶1 和 5∶1)复接成 4 路速度为 25.781 25 GHz 的 1 比特输出数据。

本文研究的第 1 种变速箱是一种基于两级移位寄存器结构的 66∶8 变速箱，如图 2[14]所示。该变速箱主要包括 4 个模块：2 个 66 比特寄存器 gear_reg1 和 gear_reg2(组成 1 个 2 级 66 比特的移

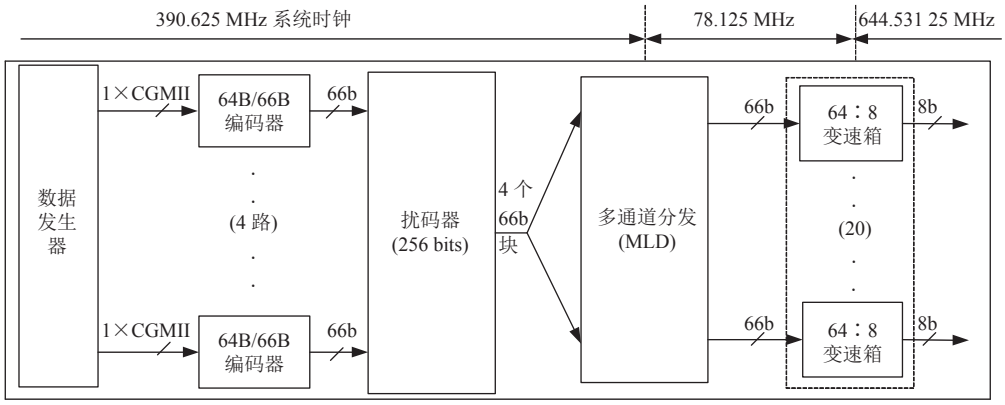


图 1　100GE 发送端物理编码子层结构框图

Fig. 1　Block diagram of 100GE PCS

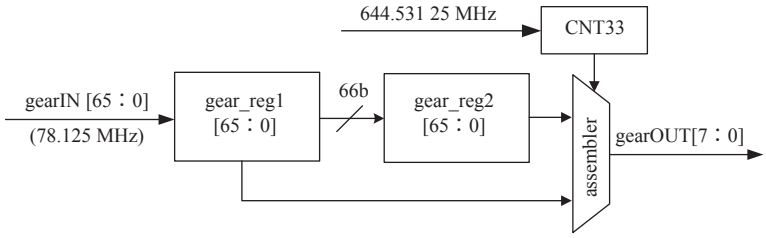


图 2　基于两级移位寄存器 66：8 变速结构框图[14]

Fig. 2　66：8 gearbox based on 2-stage shift registers [14]

位寄存器)、1 个模 33 计数器 CNT33 和 1 个拼接器 assembler。图 3 是该变速箱的时序图。

66 比特输入数据以 78.125 MHz 的速度依次进入两级移位寄存器 gear_reg1 和 gear_reg2 中，输出数据在 CNT33 的控制下通过拼接器 assembler 按一定的规律从两级移位寄存器取出。具体地，每一个输入数据首先进入第 1 个寄存器 gear_reg1，然后在下一个时钟进入第 2 个寄存器 gear_reg2。从图 3 可以看出，当 CNT33 等于‘8’时，第 1 个输入数据已经稳定地存入到第 1 个寄存器 gear_reg1 中，此时拼接模块可以开始取第 1 个输出值，输出时钟频率为 644.531 25 MHz。其中，每 4 个输入时钟或 33 个输出时钟完成一个回合(round)的读写操作，即 4 个 66 比特的输入数据被转换成 33 个 8 比特的输出数据(4×66＝33×8)。顺便说一下，其他 3 种变速箱一个回合的读写操作也是如此。事实上，这种变速箱取第 1 个输出值时，CNT33 的范围可以从‘8’到‘15’。如果在 CNT33 等于‘8’时开始取第 1 个输出值，那么在一个回合内，除了 3 个拼接时刻，输出结果都来自第 1 个寄存器 gear_reg1。如果在 CNT33 为‘15’时开始取第 1 个输出值，那么在一个回合内，除了 3 个拼接时刻，输出结果都来自第 2 个寄存器 gear_reg2。例如第一种情况，即 CNT33 等于‘8’时开始取第 1 个输出值，第一个拼接点是 CNT33 等于‘16’，此时输出数据为‘{gear_reg1[5：0], gear_reg2[65, 64]}’。另外两个拼接点是 CNT33 等于‘24’和‘32’，它们相应的输出数据分别是‘{gear_reg1[4：0], gear_reg2 [65, 62]}’和‘{gear_reg1[1：0], gear_reg2 [65, 60]}’。这种结构的变速箱的缺点是：一是拼接点多(共有 3 个)，要知道拼接点是数据最容易出错的地方，故拼接点多意味着出错的机率就越大；二是数据存入寄存器后，不管是

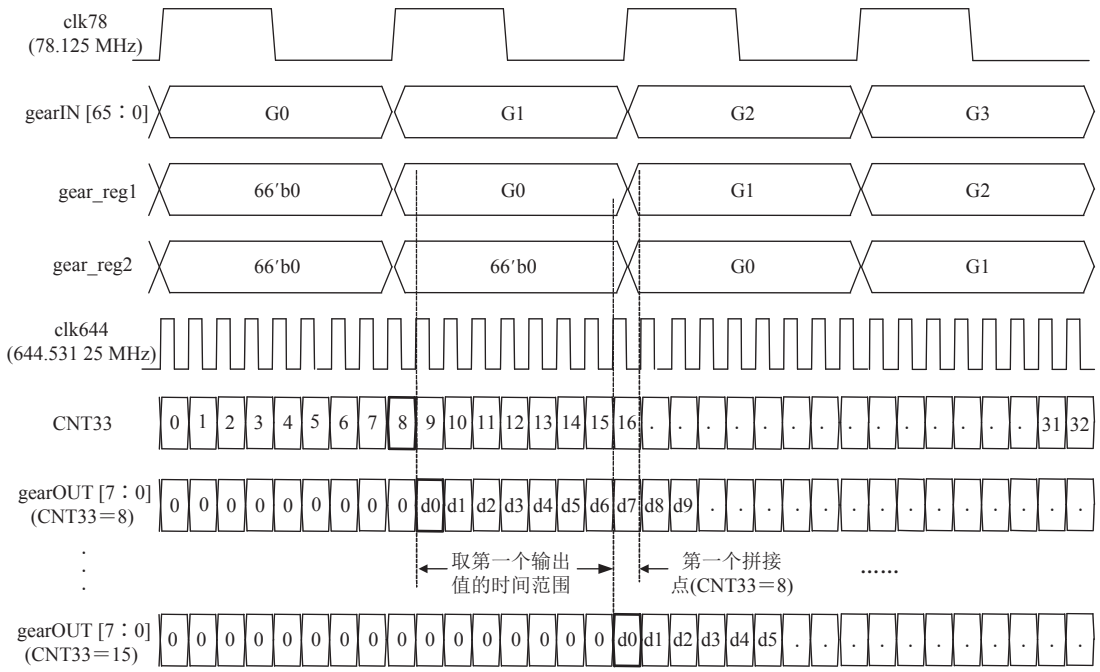


图 3 第 1 种变速箱时序图

Fig. 3 Timing diagram of the first gearbox

存入第 1 个还是第 2 个寄存器，只保留了一个输入时钟周期的时间。虽然取第 1 个输出值可以在 CNT33 的一段范围内进行，但一旦开始取输出值，无论此时 CNT33 等于‘8’到‘15’中的哪一个值，每一个输出数据都必须在某一具体的时刻取出。因此，该变速箱极易受输入输出时钟相位差的影响。这两个缺点导致这种结构的变速箱性能很差，速度不高而且容易出错。

3 基于普通寄存器的 66：8 变速箱

第 2 种变速箱是一种基于普通寄存器结构的 66：8 变速箱。这种结构的变速箱主要包括一个分选器 DEMUX、一个 88 比特的普通寄存器 REG[87：0]和两个计数器，即模 4 计数器 CNT4 和模 33 计数器 CNT33，具体如图 4 所示。图 5 为这种变速箱的存取过程示意图，为了便于直观理解，示意图中用 11 个 8 比特的寄存器组代替 88 比特的寄存器 REG[87：0]。在 CNT4 的控制下，每个回合的第 1 个输入数据存入 REG[65：0]，第 2 个输入数据存入{REG[43：0], REG[87：66]}，第 3 个输入数据存入{REG[21：0], REG[87：44]}，第 4 个输入数据存入 REG[87：22]，输入时钟为 78.125 MHz。输出很简单，以 644.531 25 MHz 的速度依次从 REG[87：0]取出(低位在前，每次取 8 比特)。图 6 是这种结构变速箱的时序图。从图 6 很容易看出，这种变速箱取第 1 个输出值的时间点有 3 个，分别是 CNT33 等于‘8’、‘9’和‘10’，原因是第 1 个输入数据的高 22 比特都在寄存器中保留了两个输入时钟周期。例如，G0[65：44]在 CNT4 等于‘1’和‘2’时保留在寄存器中，而 G1[65：44]是在 CNT4 等于‘2’和‘3’时保留在寄存器中。很明显这种变速箱没有拼接点，因此，它的性能相当不错，唯一的小瑕疵是取第 1 个输出值时的 CNT33 范围不宽，只有 3 个时间点。如果在 CNT33 等于‘9’开始取第 1 个输出值，则该变速箱可以克服输入

输出时钟的相位差。

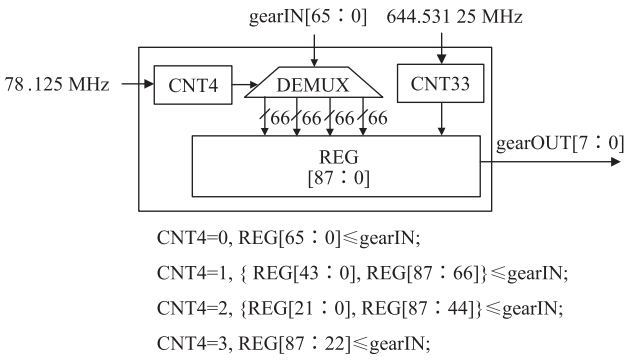


图 4　基于普通寄存器 66：8 变速箱结构框图

Fig. 4　66：8 gearbox based on ordinary register

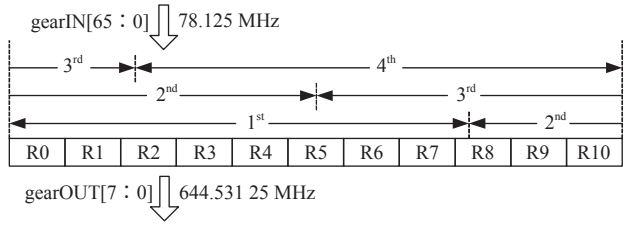


图 5　第 2 种变速箱存取示意图

Fig. 5　Access graph of the second gearbox

4　基于串入并出(SIPO)寄存器的 66：8 变速箱

第 3 种变速箱是一种基于串入并出(SIPO)寄存器结构的 66：8 变速箱。这种 SIPO 寄存器其实是先进先出(FIFO)寄存器的一种变形，具体如图 7 所示。该变速箱主要包括一个模 33 计数器，一个 264 比特的寄存器 REG[263：0]和 4 个 66 比特的寄存器 reg0、reg1、reg2 和 reg3。这 4 个 66 比特的寄存器构成了一个 4 级的移位寄存器。在该变速箱的前 4 个输入时钟，输入数据以 78.125 MHz 的速度依次存入 4 级移位寄存器 reg3、reg2、reg1 和 reg0 中。而在第 5 个输入时钟刚开始时，以输出时钟 644.531 25 MHz 的速度把这 4 个 66 比特寄存器的数据同时取出并存入 264 比特的寄存器 REG[263：0]中，紧接着在输出时钟的下一拍开始从寄存器 REG[263：0]中取输出值(低位在前，8 比特一组)。图 8 给出了该变速箱从第 5 个上升沿开始的时序图，即从

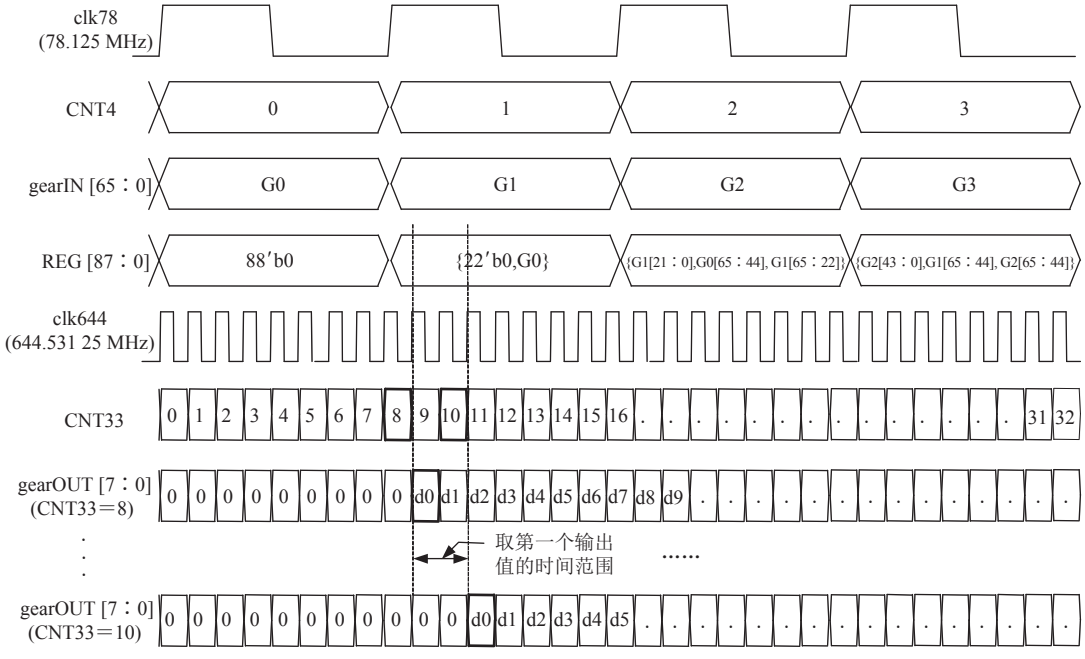


图 6　第 2 种变速箱时序图

Fig. 6　Timing diagram of the second gearbox

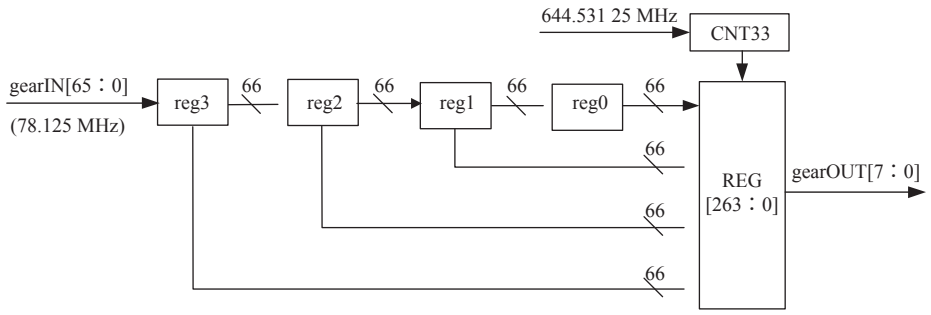


图 7　基于串入并出寄存器 66：8 变速箱结构框图

Fig. 7　66：8 gearbox based on SIPO register

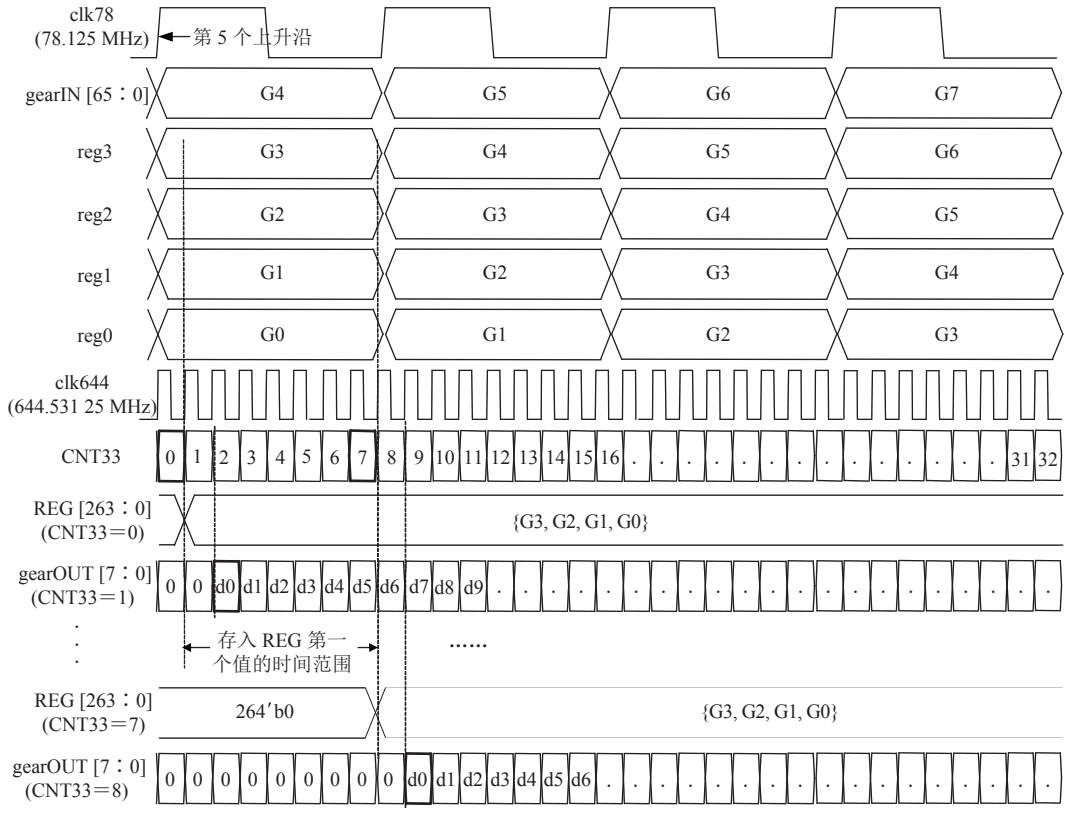


图 8　第 3 种变速箱时序图

Fig. 8　Timing diagram of the third gearbox

第 2 个回合开始给出的。从图 8 可以看出，这个 264 比特的寄存器 REG[263：0]从 4 级移位寄存器中取第一个数据的时间范围可以从 CNT33 等于‘0’到‘7’，存入寄存器 REG[263：0]的数据可以保持 4 个输入时钟周期，即一个回合的时间，也就是说在下一个回合才会改变。虽然这种结构的变速箱可以在一个时间段把 4 级移位寄存器中的值传给寄存器 REG[263：0]，但从寄存器 REG[263：0]取第 1 个输出值却必须在数据存入 REG[263：0]的第二拍(输出时钟的节拍)开始，否则，输出结果 100% 出错。这种结构的变速箱的优点是没有拼接点，缺点是输出必须在某一时间点开始，也就意味着极易受输入输出时钟相位差的影响。

5　基于轮循存储方式寄存器 66∶8 变速箱

第 4 种变速箱是一种基于轮循存储方式寄存器结构的变速箱，具体如图 9[14]所示。该变速箱主要包括一个分选器 DEMUX，一个 132 比特的寄存器 REG[131∶0]（分成两个寄存器 REG[131∶66]和 REG[65∶0]），一个拼接器 assembler 和两个计数器（即模 2 计数器 CNT2 和模 33 计数器 CNT33）。

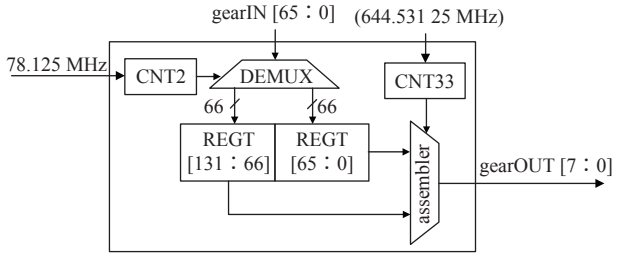


图 9　基于轮循存储方式寄存器 66∶8 变速箱结构框图[14]

Fig. 9　66∶8 gearbox based round-robin saving way register [14]

66 比特的输入数据以 78.125 MHz 的速度进入分选器 DEMUX 中。如果 CNT2 等于‘0’，则输入的数据存入 REG[65∶0]中，而 REG[131∶66]保持不变；如果 CNT2 等于‘1’，输入的数据则存入 REG[131∶66]中，而 REG[65∶0]保持不变。换句话说，每一个 66 比特的数据都可以保留两个输入时钟周期。从该变速箱的时序图中很容易看出，G0 和 G1 都在寄存器 REG[131∶0]中保留了两个输入时钟周期，具体如图 10 所示。因此，取输出数据的时间相当充裕。输出数据的取出方式是当输入数据稳定存入寄存器 REG[131∶0]中之后，以 644.531 25 MHz 的速度依次从寄存器 REG[131∶0]取出（低位在前，每 8 比特一组）。而当 REG[131∶0]取过两次值后，完成一个回合。从图 10 可以看出，这种结构的变速箱取第 1 个输出值的时间范围可以从 CNT33 等于‘8’到‘15’，且该变速箱每个回合只有一个虚拟的拼接点。例如，如果在 CNT33 等于‘8’开始取第一个输出值，唯一的虚拟拼接点是当 CNT33 等于‘24’时，对应的输出值为‘{REG[3∶0], REG[131∶128]}。为什么说是虚拟的拼接点呢？这是因为，如果把 REG[131∶0]看作一个

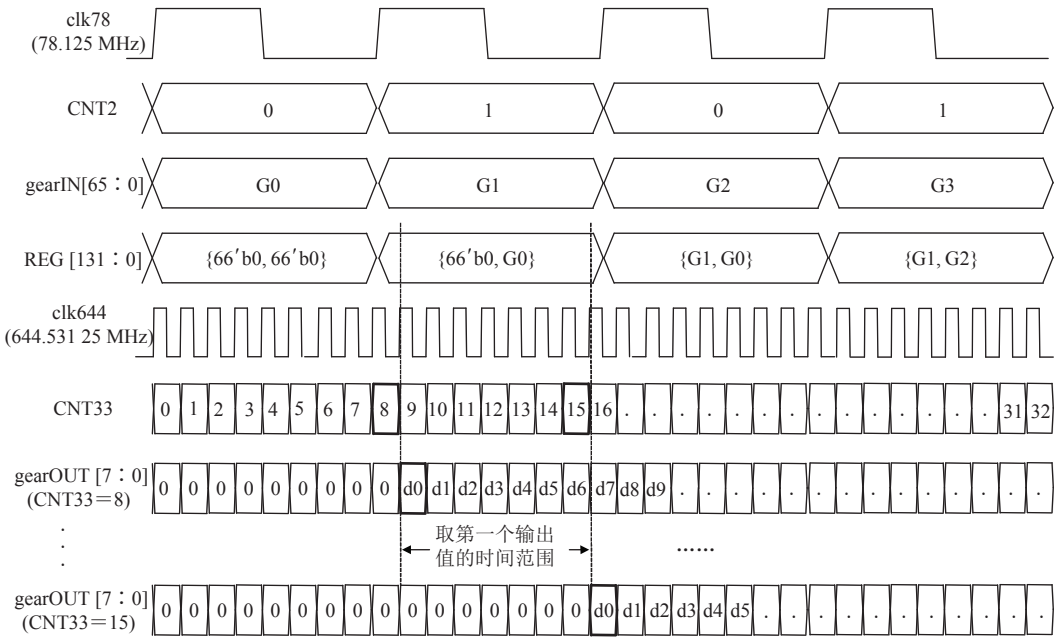


图 10　第 4 种变速箱时序图

Fig. 10　Timing diagram of the fourth gearbox

首尾相连的环形寄存器，那么这个虚拟的拼接点就不复存在了。这种结构的变速箱优点明显，近似没有拼接点(只有一个虚拟的拼接点)，如果在CNT33 等于‘9’到‘14’范围中的任一时刻开始取第 1 个输出值，则输出结果不受输入输出时钟相位差的影响，也可以说它是一种与相位无关的结构。

6 电路综合及仿真和测试

图 11 为第 4 种变速箱的功能仿真图。由于电路采用了流水线结构，输入信号与输出信号相差的节拍比较多，为了便于对照比较，将该图拆分成(a)和(b)两张图。从图 11 可以看出，在时钟的前 8 拍，第一个 66 位数据“3 c37a 5cf8 3c85 a305”(十六进制)以最低有效位(LSB)导前的方式 8 位一组依次输出，即“05”、“a3”、“85”、“3c”、“f8”、“5c”、“7a”、“c3”；在第 9 拍时，第二个数据的最低 6 位“001101”(二进制)与第一个数据的最高两位“11”(二进制)拼接成一个 8 位数据“00110111”(二进制)，即“37”(十六进制)，然后第二个数据剩下部分以 LSB 导前的方式 8 位一组依次输出。

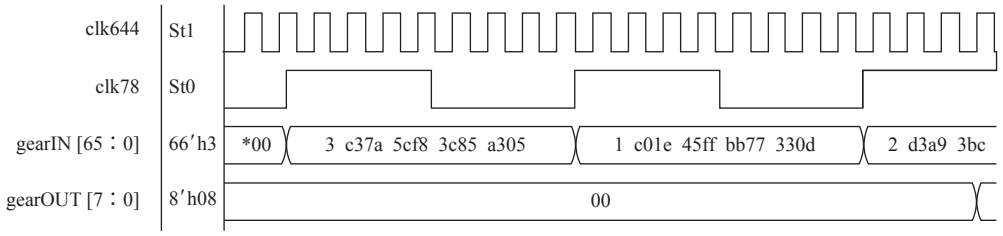
表 1 为 4 种变速箱采用 0.18 μm CMOS 标准单元库进行综合后的结果比较图(都进行了优化以及采用了流水线结构)。其中，综合工具为 Synopsys 公司的 Design Compiler。从表 1 可以看出，只有第 2 种和第 4 种变速箱满足设计要求，速度均超过设计指标 644.531 25 MHz。虽然这两种结构的变速箱的功耗和单元个数非常接近，但第 4 种具有更高的稳定性，故本文最终选择第 4 种变速箱应用到 100GE 发送端 PCS 电路中。

表 1 四种变速箱综合结果比较

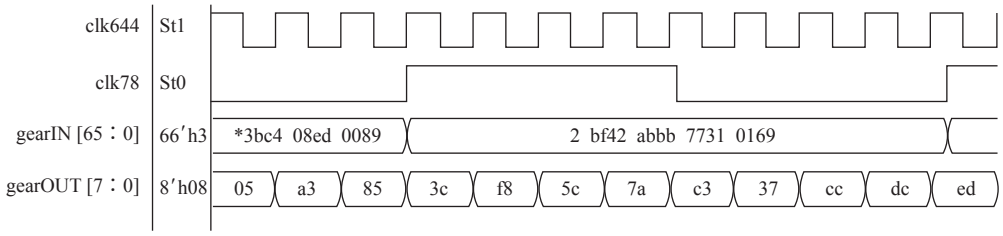
Table 1 Comparison of four gearboxes synthesis results

变速箱结构	功耗(mW)	Clk644(MHz)	单元数	稳定性
1	2.41	384.6	239	差
2	14.83	684.9	1 150	好
3	22.53	625.0	1 709	好
4	18.47	704.2	1 065	最好

含有第 4 种变速箱的 100GE 发送端 PCS 电路采用 0.18 μm CMOS 工艺流片。图 12 为整个 PCS 电路的芯片照片，它的面积为 2.89 mm^2(包



(a) 仿真输入的数据



(b) 仿真输出的数据

图 11 第 4 种变速箱的功能仿真图

Fig. 11 Function simulation graph of fourth gearbox

括焊盘)[14]。

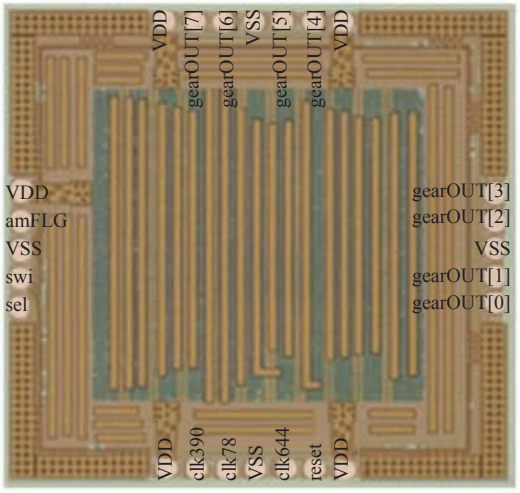


图 12　100GE 发送端 PCS 电路芯片图

Fig. 12　Chip photograph of 100GE PCS

图 13 和图 14 分别给出了该 PCS 电路第一路输出的功能仿真和测试结果(测试时钟频率为 644.531 25 MHz)，包括的信号有 amFLG 和 gearOUT[2：0](从上到下)。其中，amFLG 为对齐码的指示信号；gearOUT[2：0]为第一路输出的第 4～6 位。从图 13 可以看出，当 amFLG 为‘1’时，输出信号 gearOUT[2：0]

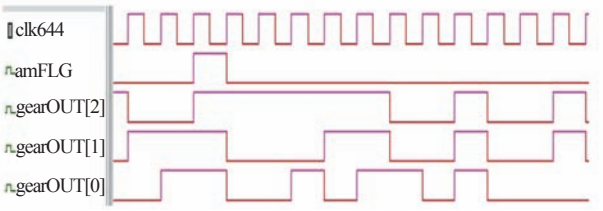


图 13　100GE 发送端 PCS 电路仿真结果

Fig. 13　Function simulation graph of 100GE PCS

图 14　100GE 发送端 PCS 电路测试结果

Fig. 14　Measured results of 100GE PCS

为“111”，然后依次是“100”、“100”、“101”、“110”、“111”、“001”、…。从图 14 可以很明显地看出，测试结果与仿真结果一致，说明该 PCS 电路可以在 644.531 25 MHz 时钟下可靠而稳定地工作，即可以达到 100 Gb/s 的传送速率，从而也证实了变速箱设计的正确性。

7 结 论

以具体比例 66∶8 变速箱为模型，采用 4 种不同的设计方法对该变速箱进行设计，并对它们进行时序分析、了解它们的优缺点及进一步比较它们的综合结果。最终从单元数量(面积)、功耗、速度和稳定性进行综合考虑，选出其中最优的一种应用到 100GE 发送端 PCS 电路中，且该 PCS 电路中的 66∶8 变速箱设计时钟高达 644.531 25 MHz。这种最优变速箱是一种基于轮循存储方式的寄存器结构，其优点是不受输入输出时钟相位差的影响，即与相位无关，从而大大地提高电路的速度和稳定性。经过结构的优化以及采用流水线结构设计，最后该变速箱的时钟速度超过 700 MHz，满足设计要求。采用 0.18 μm CMOS 工艺对该 PCS 电路流片，测试结果表明该电路能够以 100 Gb/s 的速率稳定工作。当然这种与相位无关结构也适用于其他比例的变速箱，如 33∶8 变速箱、33∶32 变速箱等。

参 考 文 献

[1] IEEE. IEEE P802.3ba 40 Gb/s and 100 Gb/s ethernet task force [OL]. 2010-07-19[2019-07-20]. http://www.ieee802.org/3/ba/. 2010.

[2] Nicholl G, Gustlin M, Trainin O. A physical coding sublayer for 100GbE [J]. IEEE Communications Magazine, 2007, 45(12): 4-10.

[3] Cole C, Allouche D, Flens F, et al. 100GbE-optical LAN technologies [J]. IEEE Communications Magazine, 2007, 45(12): 12-19.

[4] Daikoku M, Morita I, Taga H, et al. 100 Gb/s DQPSK transmission experiment without OTDM for 100G ethernet transport [J]. Journal of Lightwave Technology, 2007, 25(1): 139-145.

[5] Toyoda H, Ono G, Nishimura S, et al. 100GbE PHY and MAC layer implementations [J]. IEEE Communications Magazine, 2010, 48(3): S41-S47.

[6] Zazo JF, Lopez-Buedo S, Sutter G, et al. Automated synthesis of FPGA-based packet filters for 100 Gbps network monitoring applications [C] // 2016 International Conference on ReConFigurable Computing and FPGAs (ReConFig), 2016: 1-6.

[7] Zazo JF, Lopez-Buedo S, Ruiz M, et al. A single-FPGA architecture for detecting heavy hitters in 100 Gbit/s ethernet links [C] // 2017 International Conference on ReConFigurable Computing and FPGAs (ReConFig), 2017: 1-6.

[8] D’Ambrosia J. 100 gigabit ethernet and beyond [J]. IEEE Communications Magazine, 2010, 48(3): S6-S13.

[9] Drolet P, Duplessis L. 100G ethernet and OTU4 testing challenges: from the lab to the field [J]. IEEE Communications Magazine, 2010, 48(7): 78-82.

[10] Xia TJJ, Wellbrock G. 100G technology development for optical transport networks [C] // The 16th Opto-Electronics and Communications Conference, 2011: 395-396.

[11] Tomizawa M. 100G DWDM transport systems: driving the technologies and deployment [C] // The 10th International Conference on Optical Internet (COIN2012), 2012: 30-31.

[12] Ochi H, Sasaki S. Feasibility study of 100G ethernet with carrierless amplitude and phase modulation [C] // 2016 21st OptoElectronics and Communications Conference (OECC) Held Jointly with 2016 International Conference on Photonics in Switching (PS), 2016: 1-3.

[13] Tzimpragos G, Kachris C, Djordjevic IB, et al. A survey on FEC codes for 100G and beyond optical networks [J]. IEEE Communications Surveys & Tutorials, 2016, 18(1): 209-221.

[14] Ruan WH, Hu QS. High speed and reliability gearbox for 100GE physical coding sublayer [J]. Electronics Letters, 2016, 52(11): 908-909.