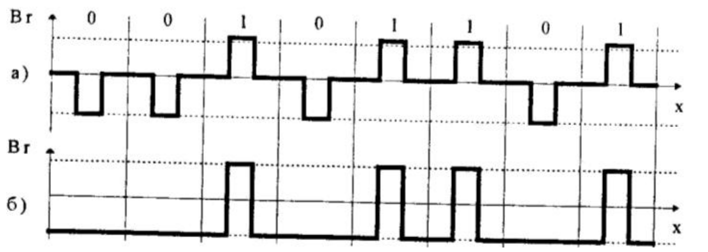
3.3 Основни методи за кодиране на информацията.

Кодовата дума е поредица от нули и единици. За да се запишат и прочетат тези “1” и “0” трябва да се кодират като сигнали. Кодирането на информация е основна функция на адаптерите при предаване на информация на разстояние по канал за връзка и на контролерите.на ВЗУ.

Методите за кодиране се делят основно на:

- **импулсни**, при които сигналът се следи по ниво;
- **потенциални**, при които сигналът се следи по фронт.

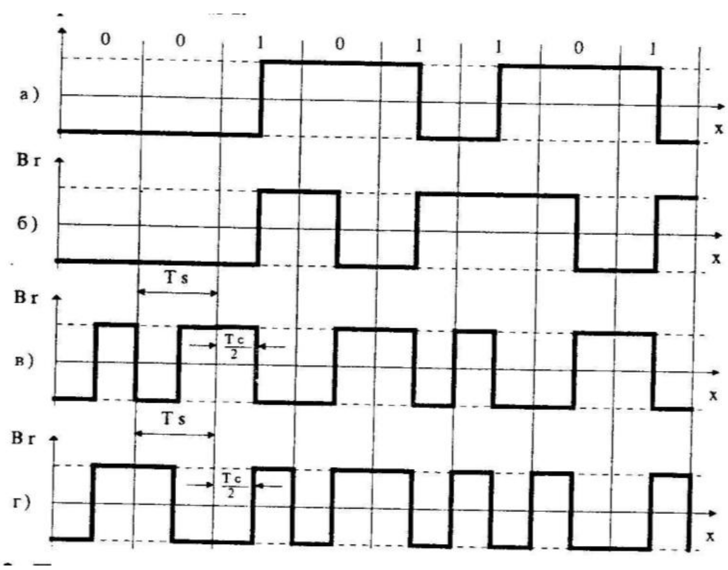
При **импулсните методи** намагнитеността на всеки участък отговаря на записаната двоична цифра. Основен недостатък на тези методи е ниската плътност на записа, поради което намират много ограничено приложение. Примери за импулсни методи:



А/ с използване на трите нива +Br, 0, - Br. Предимство: обособените участъци с противоположна намагнитеност. Недостатък: необходимост от изтриване на информацията преди запис.

Б/ с използване на две нива. Предимство: не е необходимо изтриване на стария запис. Недостатък: трудности с определяне на мястото на “0”.

Потенциалните методи са по- разнообразни и се използват широко. Те осигуряват висока плътност на запис. Примери за 4 основни потенциални метода:



А/ потенциален метод по 2 нива без връщане към нулата, **NRZ** /Non Return Zero/ (метод с реакция на “1”). При него промяната на състоянието на носителя става само при “1”.

“1”-изменение на състоянието в противоположно;

“0”-състоянието не се изменя.

Предимство: реализира се с прости схеми. Недостатък: погрешното инвертиране на един разред води до инвертиране на цялата поредица след него.

Б/ модифициран NRZ метод**, NRZI**, при който **състоянието на носителя се променя само при промяна на последователността от двоични цифри.** Предимство: прости схеми за реализиране. Недостатъци: неопределеност на първоначалното състояние; погрешното инвертиране на един разред води до инвертиране на цялата поредица**.** NRZI метода не се използва в чист вид, а е основа за методите с фазова и честотна модулация, които дават възможност да се намалят фазовите и честотни изкривявания, дължащи се на крайната дължина на отпечатъците.

В/ потенциален метод с фазова модулация, при който се използва следното съответствие:

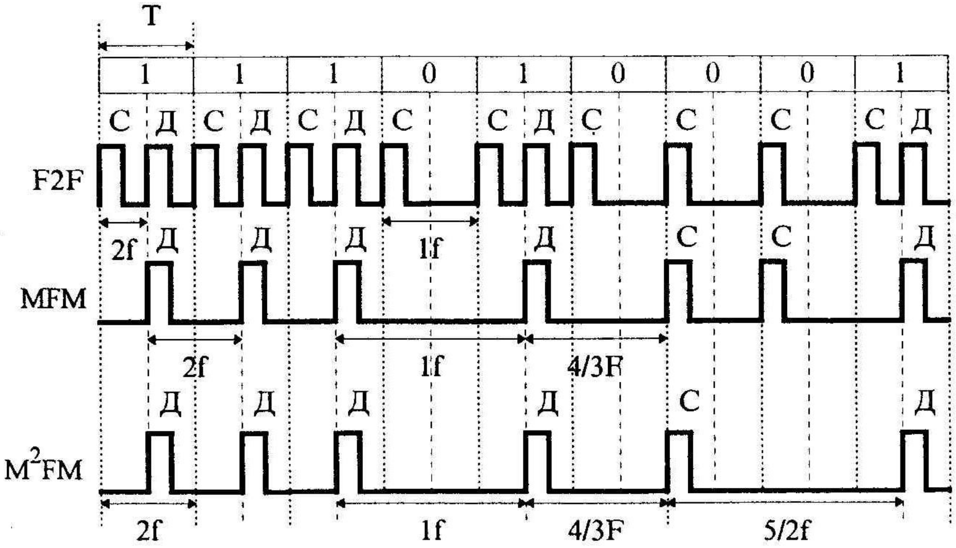
“1” – изменение на състоянието на носителя в едната посока /от “+Br” в “- Br”/;

“0” - изменение на състоянието на носителя в обратна посока /от “-Br” в “+Br”/.

При този метод при запис на последователности от еднакви цифри се извършва допълнително изменение на магнитното състояние. Методът позволява четене при препокриване на магнитните отпечатъци т.е. по-висока плътност на записа. Използва се широко, въпреки по-сложните схеми за реализирането му.

Г/ **потенциален метод с честотна модулация**, известен още като **F2F метод**. При него **състоянието на носителя се променя с една честота /F/ при кодиране на “0” и с двойно по-голяма честота /2F/ при кодиране на “1”**. Характеристиките му са идентични с тези на метода с фазова модулация, но е по-прост за реализиране.

Сега при ВЗУ на магнитни и оптични дискове най-често се използват модифицирани потенциални методи с честотна модулация. Причина за това освен осигуряването на висока плътност е и възможността за самосинхронизация. Това означава, че при предаване на информацията се използват допълнителни, синхронизиращи импулси за разграничаване на отделните битове. Те осигуряват достоверност и сигурност при предаването и четенето на информацията.



При **метода F2F** се предава една основна поредица от импулси, синхроимпулси /С/ за всеки бит информация. Между два съседни синхроимпулси се предава един бит от информацията, като за "1" има допълнителен импулс, т.н. импулс данни, "Д" между два синхроимпулса, а за "0" няма такъв импулс. Така при кодиране на "0" периодът е "T", а честотата съответно "F", а при кодиране на "1" – "T/2", и съответно - "2F".

При **модифицираната честотна модулация или код на Милер /MFM/** за синхронизация се използват и импулсите за данни. **Общото правило** при кодиране по метода **MFM** е:

- **предаване на импулса за данни "Д" за всяка единица от кодовата дума;**
- **предаване на синхроимпулса за всяка втора и всяка следваща "0" от група последователни "0" в кодовата дума.**

При това се получават 3 честоти:

- **период "T", честота "F" при информация "101";**
- **период "3T/4", честота "4 F/3" при информация "100";**
- **период "T/2", честота "2F" при информация, която се състои само от "0" или само от "1".**

MFM метода дава възможност за реализиране на запис с два пъти по-голяма плътност от F2F.

4. Контрол и корекция на информацията.

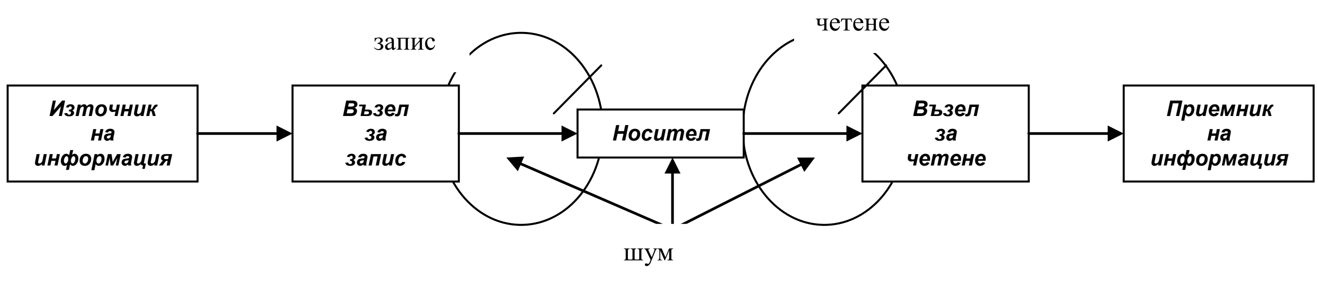
4.1. Основни принципи.

При четенето, записа и предаването на информация има въздействия /шумове/, които изкривяват и дори унищожават информацията. Шумовете са различни по произход и начин на въздействие: токови удари, електромагнитни въздействия, дефекти в носителя и др. **Шумовете** могат да се класифицират в две групи:

- **детерминирани / определени**/. Отстраняват се чрез методи, които се залагат в апаратурите за запис и четене т.е. хардуерно;
- **случайни**. Появяването им има вероятностен характер. По принцип са неотстраними хардуерно. За отстраняването им се използва специално кодиране на информацията.

Проблемът за защита от случайни грешки е най-голям при ВЗУ, поради големите плътност на информацията и скорост на запис и четене и при предаване на информацията на разстояние. Поради това при периферните устройства въпросът за кодиране на информацията е тясно свързан с този за контрола и корекцията й.

Обобщена схема за запис и четене.



Шумовете са разпространени по целия тракт от устройството от източника на информацията, до приемника, но върху схемата за удобство те са съсредоточени само в носителя и самите процеси на запис и четене.

Информацията, която се предава по канала за връзка, е двоична и **може да бъде описана с т.н. вектор W**, който се характеризира **с определена дължина L**. В най-прости случаи се използват едномерни двоични вектори.

Нека **информационният вектор W** има значение **W=01011011001**

В случая дължината L е равна на броя на разредите – **L=11**.

От друга страна шумът, водещ до грешки, може да се представи чрез друг, **шумов вектор *E***, който се характеризира със същата дължина, както информационния (приема се, че действа върху цялата кодова дума). Въздействието му върху кодовия вектор може да се даде чрез операцията сумиране по модул 2. Вследствие въздействието му, ще се получи един **изкривен вектор W*,** както е показано в примера:

$$
\begin{array}{l}
W \;\; = 01011011001 \\
\;\;\oplus \\
E \;\;\; = 0\underline{\mathbf{1}00\mathbf{11}}00000 \\
\;\;\; \underline{b=5} \\
W^* = 0\mathbf{0}010\mathbf{01}11001
\end{array}
$$

От този пример се вижда, как **трите единици в шумовия вектор Е водят до три инвертирания на кодовата дума** (показани по-плътно). Възникнала е тройна или трикратна грешка или е възникнал пакет от грешки с дължина $b = 5$.

Кратността на грешката *t* се определя от броя единици в шумовия вектор Е, които предизвикват съответния брой инвертирания в получения вектор W^* (в случая $t=3$).

Пакетът от грешки, **b** е ограничен от двата крайни ненулеви компонента на шумовия вектор (b=5). Повечето от грешките се получават във вид на пакети при работа с външни памети на магнитни и оптични носители.

Откриването на грешки се гради върху идеята, от всички кодови комбинации в даден код, No, да се използват само една част, N<No. Използваните кодови комбинации се наричат разрешени, а неизползваните – забранени. Грешката се открива ако в резултат на шумовото въздействие от разрешена кодова комбинация се получи забранена.

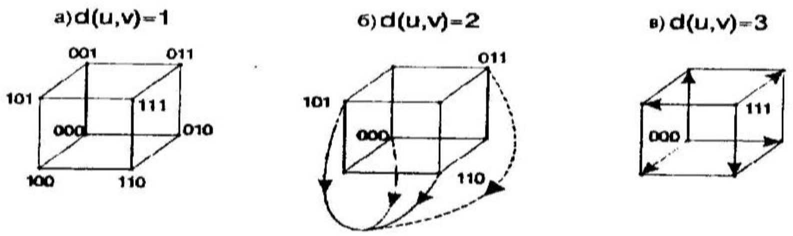
Следователно всеки код с излишък (N<No) дава възможност за откриване на грешки, като вероятността за откриване е право пропорционална на излишъка.

Важна характеристика на всеки код е т.н. **кодово разстояние, D,** което се изразява като различие между две кодови думи.

Най-разпространената мярка за разстояние между кодови думи е **кодово разстояние по Хеминг**, което се дефинира като **разлика в съдържанието на разредите на две кодови думи** – $D(u, v)$. Нека $u=11010$, а $v=\mathbf{10001}$ – вижда се, че разлика има в три разреда т. е. $D(u, v)=3$.

Най-малката разлика в разредите на две кодови думи (два кодови вектора) за цялото кодово пространство се дефинира като минимално кодово разстояние по Хеминг – D_{min}. Този параметър показва ефикасността на кода за откриване и коригиране на грешки.

Примери: Един от начините за представяне на кодовите думи е обемния, при който на всяка дума се съпоставя точка от "n" мерното пространство. Кодовите думи са разположени във върховете на куб.



А) Dmin=1. Този код няма никаква възможност за откриване на грешка, защото при всяка грешка ще се попадне в друга разрешена комбинация.

X	0	0	0	0	1	1	1	1
Y	0	0	1	1	0	0	1	1
Z	0	1	0	1	0	1	0	1

Б) Dmin=2. От възможните 8 кодови думи са използвани 4, подбрани така, че кодовото разстояние между две кои да е от тях да бъде 2. Този код има възможност за откриване на еднократна грешка, но не и за корекция.

X	0	0	1	1
Y	0	1	0	1
Z	0	1	1	0

Б) Dmin=3. От всички кодови комбинации са избрани 2. Съществува възможност за откриване на двукратна грешка и корекция на еднократна.

Условието да бъде коригирана *t*-кратна грешка се дава чрез израза: $D_{min} \geq 2t+1$,
където t е кратността на грешката при корекция.

Условието да бъде гарантирано откриването на t-кратна грешка се дава чрез израза:

$$D_{min} \geq t+1,$$

където t е кратността на грешката при контрол.

Гарантирано откриване на грешка означава, че се дава 100% гаранция, че такава грешка ще бъде открита. За по-голяма кратност на грешката няма гаранция – някои грешки могат да бъдат открити, други – не.

Кодове за откриване на грешка, изискващи предаването на информацията по канала за връзка да бъде повторено, се наричат **контролиращи кодове**. Процесът по откриване на грешката и повторно предаване на информацията по канала за връзка в случай на грешка се нарича **контрол на информацията**.

Кодове, чрез които се коригира грешка се наричат **коригиращи кодове**. Това в общия случай се извършва без да е необходимо повторно предаване на информацията по канала за връзка. Процесът по откриване и корекция на грешката се нарича **корекция на информацията**. Тези кодове изискват по-голям излишък и се използват там, където средата е силно зашумена (малка е вероятността да бъде прочетена правилно информацията) и където се търси по-голямо бързодействие (няма време за повторно предаване).

Коригиращи кодове се използват в твърдите магнитни дискове, оптичните дискове и др. Контролиращи кодове се използват при гъвкавите магнитни дискове, в повечето случаи на предаване на разстояние на информация (например чрез модеми) и др.

Контролиращите и коригиращите кодове са като правило равномерни и една от основните им характеристики е дължината на кодовата дума l.

Под $(q\text{-}l)$ контролиращ код се разбира код, който при дължина на кодовата дума l дава възможност да се откриват всички грешки с кратност от 1 до q във всяка от разрешените кодови думи.

Под $(p\text{-}l)$ коригиращ код се разбира код, който позволява в процеса на декодиране да се отстраняват възможните грешки с кратност до p във всяка от разрешените кодови думи при дължина на кодовата дума l.

Възможно е един код да бъде $(q\text{-}l)$ контролиращ и/или $(p\text{-}l)$ коригиращ: $(p \wedge q\text{-}l), (q \vee p\text{-}l)$.

Всички известни коригиращи и контролиращи кодове се делят на две групи:

- **разделими**, кодовата дума на които се дели на две полета (части) – информационно и контролно, които могат да се отделят. Информационното поле се използва за непосредствено кодиране на записваните знаци и ако липсват шумове, е напълно достатъчно за възстановяване в процеса на декодиране на записания знак. Дължината на информационното поле n непосредствено определя максималния брой знаци, които могат да се кодират с избрания код (2^n).

Броят на позициите в контролното поле при разделимите кодове е $\kappa = l - n$.

Задачата за построяване на разделим контролиращ или коригиращ код се свежда до определяне на правилата, в съответствие с които към информационното поле M_i на i-тата кодова дума се присъединява съответното й контролно поле K_i:

$$K_i = f_1(M_i)$$

- **неразделими**

Важна характеристика на контролиращите и коригиращите кодове е наборът от операции над кодовите думи, спрямо които горната функционална зависимост е инвариантна.

Би могло със значително основание да се очаква, че колкото по-голям е излишъкът на един код, толкова по-големи са неговите контролни и коригиращи възможности. Излишъкът при разделимите кодове може да се изрази по следния начин:

$$R = \frac{l}{n} = \frac{n+k}{n} = 1 + \frac{k}{n}.$$

Излишъкът характеризира цената, която се заплаща за възможността да се откриват и/или поправят грешки с даден код и понякога се изразява и по следния начин:

$$R' = \frac{k}{n}.100[\%].$$

4.2. Някои контролиращи кодове.

4.2.1. Контрол по четност и нечетност.

При контрола по четност в процеса на кодиране се формира един контролен разред, който представлява сума по модул две на разредите на думата, която кодираме. По канала за връзка се предават думи, броят на единиците в които е четен. Ако бъде открита дума с нечетен брой "1", това означава, че има грешка и информацията трябва да бъде повторена. Този код е с **Dmin=2** т.е. винаги може да се открива еднократна грешка.

Пример: Кодовата дума е 8-разредна – W=10011011

Кодиране. Определяме контролния разред чрез сума по модел 2 на разредите на кодовата дума:

$1 \oplus 0 \oplus 0 \oplus 1 \oplus 1 \oplus 0 \oplus 1 \oplus 1 = 1$

Получава се кодова дума с четен брой единици W' = 10011011 *1*, която се изпраща по канала за връзка. Тя съдържа осем информационни и един контролен разред.

Декодиране. Ако резултатът на сума по модел 2 на разредите на получената кодовата дума W' е "0" – не е възникнала грешка, а ако е "1"се получава сигнал за грешка и думата трябва да бъде повторена.

$1 \oplus 0 \oplus 0 \oplus 1 \oplus 1 \oplus 0 \oplus 1 \oplus 1 \oplus 1 = 0$

Контролът по нечетност се получава по аналогичен начин, но контролният разред допълва броя на единиците до нечетен. На практика контролният разред се получава чрез инверсия на разреда, получен като сума по модел 2 на разредите на кодовата дума.

Контролът по четност и нечетностсе използва най-често при обмен на информация на къси разстояния по сериен и паралелен интерфейс в ниско зашумена среда.

4.2.2. Полиномен контролиращ код.

Полиномните кодове са едни от най-ефективните за откриване и коригиране на грешки. Използват се широко в периферните устройства. Наричат се полиномни, защото за извършване на анализ, двоичните вектори се представят във вид на полиноми. Наричат се още линейни циклични кодове, защото кодовите вектори при кодиране се получават чрез циклични премествания на вектора по определен алгоритъм. Характерни черти на използваните в изчислителната техника полиномни кодове са следните:

а) тези кодове са двоични;

б) те са разделими и дават възможност за лесно отделяне на информационните позиции от кодовата дума;

в) относителният дял на контролните разреди в кодовата дума е незначителен;

г) кодирането и декодирането се осъществяват над кодовата дума последователно с помощта на преместващ регистър с линейни обратни връзки (реализират сума по модул 2).

Всяка двоична кодова дума *W* може да бъде представена по един единствен начин, с помощта на полином спрямо някаква променлива *x*. Всеки член на полинома отразява наличието на единица в двоичния вектор, а степента му – местоположението на тази единица.

Например: Кодовите думи: 10111; 10110010 се представят като полиноми съответно: $x^4 + x^2 + x + 1;\ x^7 + x^5 + x^4 + x$.

Правила за работа с полиноми:

$X^S + X^S = 0$	$- X^S = X^S$
$X^S + 0 = X^S$	$X^S * X^F = X^{S+F}$
$0 + X^S = X^S$	$X^S : X^F = X^{S-F}$
$0 + 0 = 0$	

Умножението на два полинома се извършва като всеки член на единия полином се умножи с всеки член на другия полином и след това се извърши сумиране. При делението на два полинома членовете на частното се получават последователно. Степента на всеки член на частното, умножена по най-високата степен на делителя дава съответната степен на делимото. Делението продължава докато съответната степен на делимото стане по-ниска от тази на делителя. При делението на два полинома се получава частно Q(x) и остатък R(x). Ако остатъка R(x) = 0 се казва, че делителя дели делимото без остатък. При това делимото може да се разложи и е равно на произведението на два полинома, делителя и частното.

Примери за действия с полиноми**:**

Събиране:

$$\begin{array}{r} X^4 + X^3 \quad\quad + 1 \\ X^3 + X^2 \quad + 1 \\ \underline{X^4 + X^3 \quad +X + 1} \end{array}$$

Резултат: $\mathbf{X^3 + X^2 + X + 1}$

Изваждане:

$$\begin{array}{r} X^4 + X^3 + \quad\quad 1 \\ \underline{X^3 + X^2 + 1} \end{array}$$

$\mathbf{X^4 + X^2}$

Умножение: $\mathbf{X^3 + X + 1}$ **или** $\mathbf{X^3 + X + 1}$

$$\begin{array}{r} X^3 + X + 1 \\ \text{x} \quad \underline{X + 1} \\ X^4 + X^2 + X \\ \underline{X^3 + X + 1} \\ X^4 + X^3 + X^2 + 1 \end{array} \qquad \begin{array}{r} X^3 + X + 1 \\ \underline{X + 1} \\ X^3 + X + 1 \\ \underline{X^4 + X^2 + X} \\ X^4 + X^3 + X^2 + 1 \end{array}$$

Резултат**:** $\mathbf{X^4 + X^3 + X^2 + 1}$ $\mathbf{X^4 + X^3 + X^2 + 1}$

Деление:$\mathbf{X^4 + X^3 + X^2 + 1 / X + 1 = X^3 + X + 1}$

$$\begin{array}{r} \underline{X^4 + X^3} \\ X^2 + 1 \\ \underline{X^2 + X} \\ X + 1 \\ \underline{X + 1} \\ 0 \end{array}$$

За получаване на контролиращ код при кодиране **думата, която се кодира се умножава с полином *G(x)*, наречен пораждащ (генериращ) полином. Степента на полинома *G(x)* определя излишъка (*k*= *l-n*). Оттук следва, че всяка от разрешените кодови думи на полиномния код се дели без остатък на *G(x)*. Именно тази особеност на разрешените кодови думи се използва за откриване на грешки в прочетената кодова дума *W'*.**

Декодирането **на полиномния код се свежда до разделяне на прочетената кодова дума *W'(x)* на *G(x)*. Ако делението се осъществява без остатък не е допусната грешка; в противен случай се счита, че е допусната грешка**.

Отделянето на *W* от *W'* (декодирането) се осъществява най-просто, ако кодирането се изпълни не чрез умножение, а чрез:

1/ думата, която ще се кодира, *W* се измества с "к" разреда наляво т.е. осъществява се операцията $x^k.W(x)$ – *"к"* е най-високата степен на пораждащия полином;

2/ резултатът се дели на пораждащия полином и като резултат се получава частно Q(x) и остатък R(x) / $x^k.W(x) / G(x)=Q(x)+R(x)$;

3/ остатъкът от това делене *R(x)* се прибавя към изместената дума. Контролни са последните разреди. Броят им е равен на степента на пораждащия полином.

Пример**: Думата, която ще кодираме: W = 10001011, съответния полином е: W(x)=** $\mathbf{X^7 + X^3 + X + 1}$**. Пораждащия полином е:** $\mathbf{G(x) = X^5 + X^2 + 1}$**.**

1/ Изместваме W(x) с к=5 разреда т.е. умножаваме с X^5: $X^5 * (X^7 + X^3 + X + 1) = X^{12} + X^8 + X^6 + X^5$

2/ Делим резултата на пораждащия полином: $X^{12} + X^8 + X^6 + X^5 / X^5 + X^2 + 1 = X^7 + X^4 + X^3 + X^2$

$$\begin{array}{r} \underline{X^{12} + X^9 + X^7} \\ X^9 + X^8 + X^7 + X^6 + X^5 \\ \underline{X^9 + X^6 + X^4} \\ X^8 + X^7 + X^5 + X^4 \\ \underline{X^8 + X^5 + X^3} \\ X^7 + X^4 + X^3 \\ \underline{X^7 + X^4 + X^2} \\ X^3 + X^2 \end{array}$$

$\mathbf{R(x) = X^3 + X^2}$

3/ **Получава се полинома $W'(x) = X^{12} + X^8 + X^6 + X^5 + X^3 + X^2$ и кодираната дума, която се изпраща по канала за връзка е: W' = 10001011 01100**

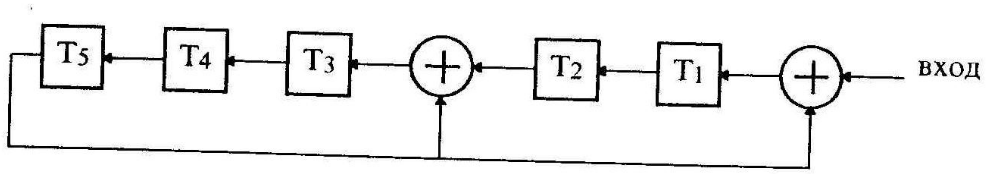
Информационни разреди Контролни разреди

При декодирането $W'(x) = X^{12} + X^8 + X^6 + X^5 + X^3 + X^2$ се дели на $G(x) = X^5 + X^2 + 1$ и ако няма грешка се получава частно и остатък $R(x) = 0$.

Най-широко приложение полиномните кодове намират при запис на информацията върху магнитни носители. Те се предпочитат пред контрола по четност /нечетност/ при попътечкова организация на записа и четенето, тъй като при една и съща вероятност за откриване на грешките използват значително по-малък брой контролни разреди.

Възможността да се използват полиномните кодове следва от простотата на кодирането и декодирането, реализирани на базата на преместващи регистри с линейна обратна връзка. Регистърът съдържа толкова запомнящи клетки, колкото е степента на пораждащия полином / в примера –5/ и толкова обратни връзки, колкото са членовете на пораждащия полином минус 1 / в примера 3-1=2/. Обратната връзка се взема след старшия разред. Връзки има към клетките, съответстващи на съдържащите се в пораждащия полином членове / в примера 2 и 0 /.

Принципна схема на регистър-делител от последователен тип за кодовата дума от примера.



$\oplus$ - двувходова схема, реализираща функцията сума по модул 2.

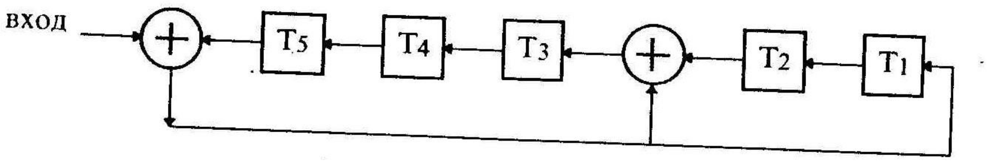
За да се извърши делението е необходимо регистъра да се нулира. След това на входа се подава последователно, такт по такт изместената с "к" разреда наляво дума т.е. $x^5.W(x)$, като се започне от най-старшия разред. Едновременно с подаването на всеки разред се извършва и циклично преместване. След "n + к" такта в регистъра остава остатъка R(X).

$x^k.W(x)$	T1	T2	T3	T4	T5
нулиране	0	0	0	0	0
1	1	0	0	0	0
0	0	1	0	0	0
0	0	0	1	0	0
0	0	0	0	1	0
1	1	0	0	0	1
0	1	1	1	0	0
1	1	1	1	1	0
1	1	1	1	1	1
0	1	1	0	1	1
0	1	1	0	0	1
0	1	1	0	0	0
0	0	1	1	0	0
0	0	0	1	1	0

$R(X) = X^3 + X^2$

Недостатък на този регистър-делител е, че делението се извършва бавно, за "n + к" такта. Бързодействието се увеличава ако се използва модифициран регистър-делител, при който входния сигнал се подава към двувходова схема, реализираща функцията сума по модул 2, която е включена след най-старшата запомняща клетка.

Модифициран регистър-делител.На входа се подава последователно W(x) като се започне от най-старшия разред. Остатъка се получава в регистъра след "n" такта.



W(x)	T1	T2	T3	T4	T5
нулиране	0	0	0	0	0
1	1	0	1	0	0
0	0	1	0	1	0
0	0	0	1	0	1
0	1	0	1	1	0
1	1	1	1	1	1
0	1	1	0	1	1
1	0	1	1	0	1
1	0	0	1	1	0

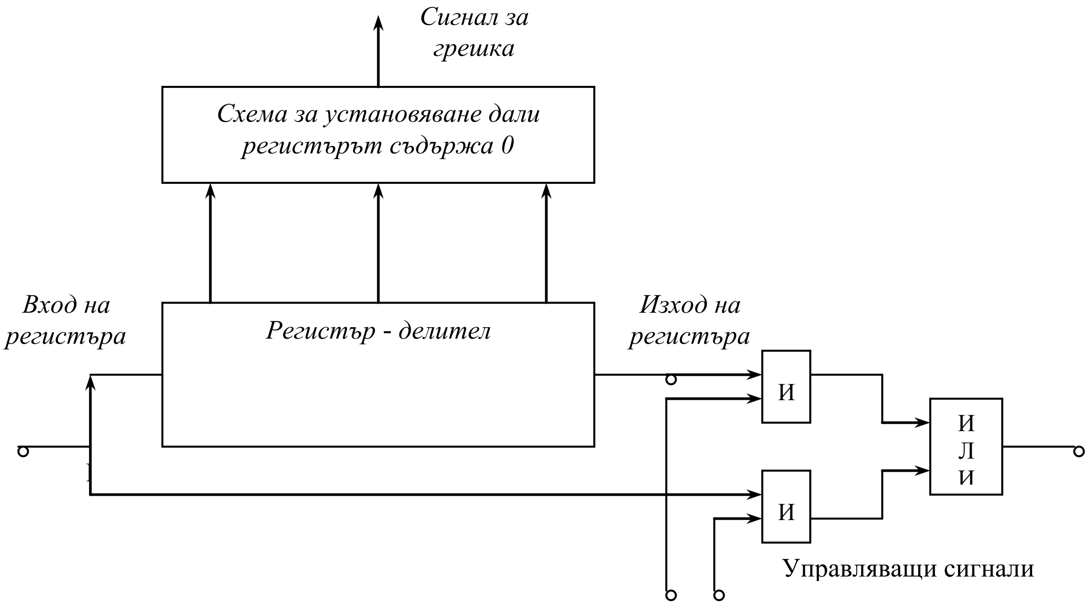
$R(X) = X^3 + X^2$

Ако четенето и записът не се извършват едновременно, един и същ регистър-делител се използва и за кодиране, и за декодиране.

Блокова схема на кодиращо-декодиращо устройство за полиномни кодове.

При кодирането подаването на *W(x)* на входа на преместващия регистър е еквивалентно на умножението му по x^k. Най-напред за *n* такта входните сигнали се подават освен към входа на регистъра и директно към изхода на схемата. При следващите *k* такта изходът на схемата се свързва към изхода на регистъра и се предават контролните разреди.

При декодиране *W'(x)* се подава към входа на регистъра и към изхода на схемата (*n* такта). След това се проверява дали съдържанието на регистъра е равно или не на нула. При втория случай се дава сигнал за грешка.



Регистър-делители от последователен тип се използват при гъвкавите магнитни дискове за получаване на контролните байтове (CRC –байтове). Регистър-делители от паралелен тип се използват при магнитни ленти. Устройствата за кодиране и декодиране се вграждат в контролерите на съответните периферни устройства.