ПРЕЦИЗІЙНІ ВИМІРЮВАННЯ ТА НОВІТНІ ТЕХНОЛОГІЇ

УДК 621.391

Н.В. ЗАХАРЧЕНКО, М.М. ГАДЖИЕВ, А.В. КОЧЕТКОВ, Е.Б. ШАМШИДИН
Одесская национальная академия связи им. А.С. Попова

ПОВЫШЕНИЕ ИНФОРМАЦИОННОЙ СКРЫТНОСТИ ПЕРЕДАЧИ НЕ РАВНОВЕРОЯТНОГО АЛФАВИТА

Предложен метод вторичного кодирования передаваемого не равновероятного первичного алфавита с целью обеспечения равновероятной передачи символа в канале с точностью до одного процента. Определено количество различных кодовых комбинаций («банк» кодовых слов символа) обеспечивающих равновероятную передачу кодовых слов. Для уменьшения времени передачи предложено использование таймерных сигналов, определена оптимальная длительность таймерного кодового слова. Показано что, с увеличением интервала реализации (m) при $S = \text{const}$ число возможных сигнальных конструкций растет. При этом растет информационная емкость одного кодового слова. Информационная емкость одного найквистового элемента при увеличении m вначале растет (до $m \le 5$), а при $m > 5$ начинает снижаться. Для каждого "m" информационная емкость с ростом "S" увеличивается. Установлено что, выбор параметра "S" целесообразно ограничивать значением $S = 8$, так как при $S > 8$ увеличивается вероятность ошибочного приема одного момента модуляции даже в "хорошем" состоянии канала. Связано это с тем, что при $S > 8$ растет значение информационной емкости одного кодового слова, но еще быстрее растут потери за счет увеличения вероятности ошибки.

Ключевые слова: информация, энтропия, интервал реализаций, таймерные сигналы, момент модуляции.

N.V. ZAKHARCHENKO, M.M. GADZHIEV, A.V. KOCHETKOV, E. B. SHAMSHIDIN
A.S. Popov Odessa national academy of telecommunications

TO INCREASE THE INFORMATION SECRECY TRANSMISSION IRREGULAR ALPHABET

The method of the secondary transfer encoding irregular primary alphabet for the purpose of ensuring uniform transmission of the symbol in the channel with an accuracy of one percent. The quantity of different code combinations ("Bank" codeword symbol) to ensure the transfer of irregular code words. To reduce the transmission time suggested the use of timing signals, determined the optimal duration timer code word. It is shown that with the increase of the interval of realization (m) when $S = \text{const}$ the number of possible signal structures increases. This increases the information capacity of a single code word. The information capacity of one niquitao element by increasing initially grows (up to $m \le 5$), and when $m > 5$ it begins to decline. For each (m) information capacity with increasing "S" increases. Established that the choice of the parameter "S" should be limited to value $S = 8$, as it increases the probability of erroneous reception of a single moment of modulation even in a "good" channel state. This is due to the fact that when $S > 8$ increasing the value of information capacity of a single codeword, but growing at an even higher loss due to increasing probability of error.

Key words: information, entropy, interval timer signals, time modulation.

При расшифровке полученного при несанкционированном доступе к передаваемому сообщению чаще всего используется статистический метод распознавания передаваемого символа, основанный на вероятностях использования отдельных символов в первичном тексте.

В работе предлагается метод кодирования не равновероятного первичного алфавита в кодовое множество с максимальной энтропией, обеспечивающего равновероятную передачу в канале.

Количество информации $I(x_i)$,содержащейся в событии x_i происходящем с вероятностью $P(x_i)$ определяется [1, 5]:

$$I(x_i) = \log_2 \frac{1}{P(x_i)} . \tag{1}$$

Для полного ансамбля событий: $X = \begin{pmatrix} x_1; & x_2; & x_3; \ldots x_n \\ P(x_1); P(x_2); P(x_3); \ldots P(x_n); \end{pmatrix}$

среднее значение $\bar{I}(x)$ информации по всему ансамблю событий [2]:

$$\bar{I}(x) = M[I(x_i)] = H , \tag{2}$$

называется энтропией сообщения (H) и измеряется в двоичных единицах на сообщение:

$$H = -\sum_{i=1}^{n} P(x_i) \log_2 [P(x_i)] . \tag{3}$$

Воспользуемся основным свойством энтропии [3]: при заданном числе "n" символов энтропия максимальна и равна:

$$H(x) = \log_2 n , \tag{4}$$

лишь тогда, когда

$$P(x_1) = P(x_2) = \ldots = P(x_n) = 1/n . \tag{5}$$

Для примера возьмем алфавит русского языка. В таблице 1 приведены $n = 32$ символа русского языка (x_i) (в том числе и символ (пр) – пробел), вероятности их появления $P(x_i)\%$ (в процентах) [6], и вероятности появления $P(x_i)\%$ округленные до ближайшего целого большего значения в процентах $E^+[P(x_i)\%]$.

С целью обеспечения в канале равновероятной передачи всех символов с точностью до одного процента предлагается для каждого подлежащего передаче символа x_i выделить количество различных кодовых комбинаций равное числу $E^+[P(x_i)\%]$.

Общее число различных комбинаций для алфавита русского языка $\sum E^+[P(x_i)\%]=116$ кодовых слов. Превышение общего числа комбинаций больше 100% является следствием округления вероятностей $P(x_i)$ до ближайшего большего целого числа процентов (табл. 1).

Все комбинации, относящиеся к конкретному символу x_i зациклены и передаются по очереди при появлении данного символа в передаваемом тексте.

Таблица 1

Вероятностные параметры символов русского языка

№ п/п	1	2	3	4	5	6	7	8	9	10	11
x_i	пр	А	Б	В	Г	Д	Е,Ё	Ж	З	И	Й
$P(x_i)\%$	17,5	6,2	1,4	3,8	1,3	2,5	7,2	0,7	1,6	6,2	1
$E^+[P(x_i)\%]$	18	7	2	4	2	3	8	1	2	7	1
№ п/п	12	13	14	15	16	17	18	19	20	21	22
x_i	К	Л	М	Н	О	П	Р	С	Т	У	Ф
$P(x_i)\%$	2,8	3,5	2,6	5,3	9	2,3	4	4,5	5,3	2,1	0,2
$E^+[P(x_i)\%]$	3	4	3	6	9	3	4	5	6	3	1
№ п/п	23	24	25	26	27	28	29	30	31	32	
x_i	Х	Ц	Ч	Ш	Щ	Ь,Ъ	Ы	Э	Ю	Я	
$P(x_i)\%$	0,9	0,4	1,2	0,6	0,3	1,4	1,6	0,3	0,6	1,8	
$E^+[P(x_i)\%]$	1	1	2	1	1	2	2	1	1	2	

$$\sum_{i=1}^{n} E^+[P(x_i)\%]=116$$

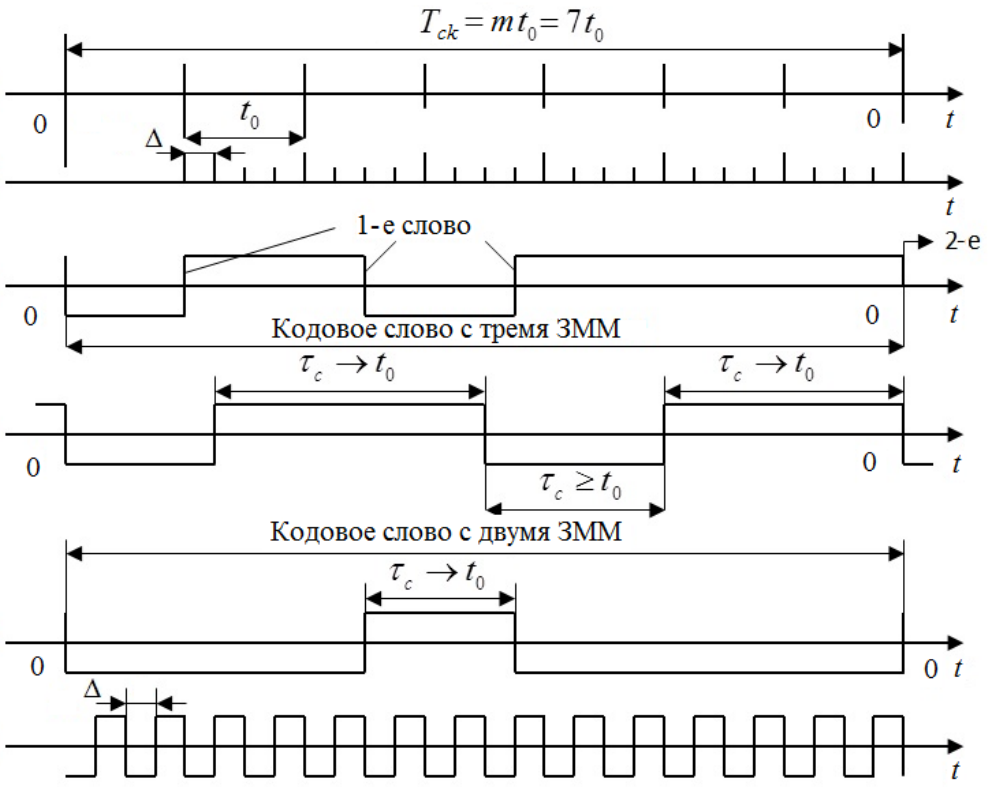

Рис. 1. Формирование сигнального алфавита бинарных ТСК на интервале времени $T_{ck}=mt_0$ при базовом элементе Δ

Таким образом, для возможности обеспечения равновероятной передачи по каналу различных кодовых конструкций, передатчик должен содержать для каждого символа x_i “банки” различных кодовых слов в

количестве $E^{+}[P(x_i)\%]$. Для алфавита русского языка таких "банков" будет 32 (по количеству передаваемых символов) с общей памятью 116 кодовых слов.

С целью экономии времени передачи воспользуемся таймерными сигналами [1]. Принцип построения таймерных сигнальных конструкций (ТСК) представлен на рис. 1, и заключается в следующем. Сигнальный алфавит бинарных ТСК формируется на интервале времени $(T_{ck} = m t_0)$, t_0 – величина найквистового элемента (обратная полосе пропускания канала), при базовом элементе отсчета длительностей Δ ($\Delta = t_0/S$, $S \in 1;2;3;...k$ – целые числа.

Таким образом, на интервале $(T_{ck} = m t_0)$ расположено $n = mS$ точек отрезков Δ.

Из всего множества 2^n возможных на интервале T_{ck} сигналов разрешенными считаются только те, в которых соседние значащие моменты модуляции (ЗММ) отстоят друг от друга на время, не меньшее чем $S\Delta = t_0$. Это условие обеспечивает минимум межсимвольных искажений. Каждый ЗММ может занимать на оси времени позиции, расположенные на расстоянии $k\Delta \geq t_0$ друг от друга, причем Δ определяется как минимальное расстояние между соседними положениями одного ЗММ в разных конструкциях. Информация о передаваемом сообщении, переносимая ТСК, содержится в номере временной позиции, занимаемой ЗММ, причем первый информационный ЗММ может появиться не раньше, чем через $S\Delta$ позиций от момента начала сигнала (нулевой позиции). Так как величина $\Delta < t_0$, то увеличение пропускной способности возможно, если число реализаций сигнала N_p на интервале (mt_0) не меньше чем 2^m. Можно показать, что при заданном $S = t_0/\Delta$ на интервале m единичных элементов мощность сигнального алфавита бинарных ТСК определяется как [4]

$$N_p(i = \text{const}) = C^i_{mS-i(S-1)}, \quad N_p(i = \text{var}) = \sum_{i=1}^{m} C^i_{mS-i(S-1)}, \tag{6}$$

$$C^i_m = \frac{m!}{i!(m-i)!}, \tag{7}$$

где i – число ЗММ в сигнальной кодовой конструкции. Среди разрешенных сигналов могут быть реализации с одним ЗММ, двумя, тремя и т. д. С максимальным числом моментов модуляции m возможна только одна реализация.

Для примера в табл. 2 приведено количество реализаций ТСК $T_{ck} = mt_0S$ и сигналов простого двоичного кода $N_p = 2^m$.

Таблица 2

Количество реализаций ТСК и среднее значение ЗММ для величин: i=3, $T_c = mt_0 \cdot S$ **;** $N_p = 2^m$

S \ m	$m = 4$, $N_p = 16$	$m = 5$, $N_p = 32$	$m = 6$, $N_p = 64$	$m = 7$, $N_p = 128$	$m = 8$, $N_p = 256$	$m = 9$, $N_p = 512$
2	10	35	84	165	286	455
3	20	84	220	455	816	1330
4	35	165	455	969	1771	2925
5	56	286	816	1771	3276	5456
6	84	455	1330	2925	5456	9139
7	120	680	2024	4495	8436	14190
8	165	969	2925	6545	12341	20825

Как видно из табл. 2, на одном и том же интервале T_{ck} можно образовать большее количество ТСК, чем сигналов простого двоичного кода $N_p = 2^m$.

Следовательно, скорость передачи, т. е. количество передаваемой информации на интервале T_{ck} увеличивается. Так как минимальное расстояние между ЗММ двух ближайших кодовых слов равно $\Delta < t_0$, а прием значащих моментов воспроизведения (ЗМВ) осуществляется методом анализа в отдельных зонах Δ [1] то, естественно, вероятность ошибочного приема такого сигнала выше, чем элемента при разрядно-цифровом коде.

Из приведенной выше информации о методе формирования ТСК на m-элементном интервале времени T_{ck} следует, что за счет значения $\Delta(S)$ одно и то же число реализаций можно получить на различных интервалах времени.

Для примера на рис. 2 представлены зависимости длительности сигнальной конструкции при заданной мощности кодового множества и параметра S. Из этих зависимостей следует, что при $S \geq 2$ для получения $N_p = 2^m$ можно затратить время $\tau_c < mt_0$. При этом это неравенство тем сильнее, чем больше значение S.

Пусть множество 116 кодовых слов представляет кодовые конструкции с тремя информационными отрезками x_1, x_2, x_3 [4]. Для возможности оценки принадлежности к передаваемому множеству потребуем, чтобы координаты x_1, x_2, x_3 удовлетворяли условию [1]:

$$x_1 + 2x_2 + 3x_3 = 0 \,(\mathrm{mod}7) \qquad (8)$$

Тогда мощность множества M из которого можно выбрать 116 кодовых слов удовлетворяющих условию (8) должна быть [4] $M \geq 116 \times 7 = 812$ к.с. Каждое из этих 812 кодовых слов должно иметь энтропию равную $\log_2 812 = 9{,}65$ двоичных единиц.

Выбрав реализации кодовых слов удовлетворяющих условию (8) на интервале $T_p = 5t_0$, мы синтезируем кодовые конструкции с максимальной энтропией [2, 4]:

$$H = \log_2 N_p \qquad (9)$$

и максимальной информационной емкостью $I_н$ найквистового элемента (t_0):

$$I_н = \frac{H}{m} = \frac{\log_2 N_p}{m}, \qquad (10)$$

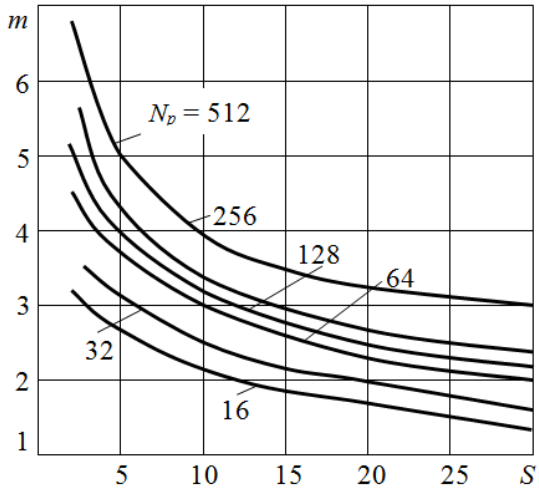

Рис. 2. Зависимости длительности сигнальной конструкции при заданной мощности кодового множества и параметра S

В таблице 2 приведены значения энтропии как функции: $H = f(m, S)$ (числитель) и информационной емкости найквистового элемента (знаменатель) для $m \in 4 \div 10$, при $S \in 2 \div 12$.

Таблица 3

Энтропия и информационная емкость найквистового элемента

S \ m	4	5	6	7	8	9	10
2	3,322/0,83	5,129/1,026	6,392/1,065	7,366/1,052	8,16/1,02	8,830/0,98	9,41/0,941
3	4,321/1,08	6,392/1,279	7,781/1,297	8,830/1,261	9,672/1,209	10,377/1,153	10,982/1,098
4	5,129/1,282	7,366/1,473	8,830/1,472	9,920/1,417	10,79/1,349	11,514/1,279	12,134/1,213
5	5,807/1,452	8,160/1,632	9,672/1,612	14,790/1,541	11,673/1,459	12,414/1,379	13,042/1,304
6	6,392/1,598	8,830/1,766	10,377/1,730	11,514/1,645	12,414/1,552	13,159/1,462	13,793/1,379
7	6,907/1,727	9,409/1,882	10,383/1,831	12,134/1,733	13,042/1,630	13,793/1,533	14,432/1,443
8	7,366/1,842	9,920/1,984	11,514/1,919	12,677/1,811	13,591/1,699	14,346/1,594	14,989/1,499
9	7,781/1,945	10,377/2,076	11,987/1,999	13,158/1,880	14,078/1,760	14,837/1,649	15,482/1,548
10	8,160/2,040	10,790/2,158	12,414/2,069	13,591/1,942	14,516/1,815	15,277/1,697	15,923/1,592
11	2,127	2,233	2,134	1,998	1,864	1,742	1,632
12	2,208	2,303	2,193	2,051	1,910	1,783	1,669

Из таблицы 3 следует:

1. С увеличением m при $S = \mathrm{const}$ энтропия реализаций также увеличивается.

2. Информационная емкость одного найквистового элемента максимальна на интервале реализации $T_p = 5t_0$ (для $S > 3$); $I_н(m = 4) < I_н(m = 5) > I_н(m = 5)$.

В качестве оптимальной длительности кодовых слов следует выбирать $T_p = 5t_0$, что соответствует максимальной информационной емкости найквистового элемента.

При передаче по каналам с гауссовским шумом следует учитывать, что вероятность ошибочного приема таймерной сигнальной кодовой конструкции – ТСК (P_o) определяется величиной зоны Δ, среднеквадратическим отклонением значащего момента восстановления (σ), которая, в свою очередь зависит от соотношения сигнал/шум, а также средним числом переходов в слове [1] ($\bar{i}$).

$$P_o = 1 - [\Phi(\Delta/2\sigma)]^{\bar{i}}, \qquad (11)$$

где $\Phi(x)$ – интеграл вероятностей:

$$\Phi(x) = \frac{1}{\sqrt{2\sigma}} \int_0^x e^{\frac{-t^2}{2}} dt. \qquad (12)$$

Учитывая, что с увеличением S растет не только число реализаций N_p, но и вероятность ошибочного приема их (P_o), то для каждого канала есть свое значение Δ_0, при котором реализуется максимальная пропускная способность системы. При этом каждая из реализаций сигнала на интервале $(T_{ck} = m t_0)$ представляет собой одну из реализаций многопозиционного во времени сигнала. Тогда значение

пропускной способности:

$$C_m = \frac{1}{m}(\log_2 N_p - H_{\text{пот}}) . \quad (13)$$

Здесь $H_{\text{пот}}$ определяет потери в канале из-за неопределенности в принятии кодовой сигнальной конструкции:

$$H_{\text{пот}} = -\left[P_{\text{п}} \log_2 P_{\text{п}} + (1 - p_{\text{п}}) \log_2 \frac{1 - P_{\text{п}}}{N - 1} \right], \quad (14)$$

где $P_{\text{п}}$ - вероятность правильного приема сигнальной конструкции с i – перехода:

$$P_{\text{п}} = 2[\Phi(\Delta/2\sigma)]^{\bar{i}} . \quad (15)$$

На рис. 3 приведены зависимости пропускной способности каналов с разным уровнем флуктуационных шумов (задано $h = (u_{\text{с}}/u_{\text{ш}})$ как функции S (кривые 1, 2, 3 для h = 7,5 и m = 8, 6, 5 соответственно, кривые 4, 5, 6 для h = 5,5 и m = 8, 6, 5 в соответствии).

Из рисунка 3 следует, что для каждого значения h является величина зоны, при которой C_m будет максимальной. На практике оптимальное значение определяется среднеквадратичным отклонением смещения фронта сигнала на выходе канала (σ_k), $\Delta_{\text{опт}}$ = (3,8 ... 5,5); σ_k = 3,8 ... 4,5/ h .

Оценка полученных результатов

Анализ таблиц 1-3 показывает:

1. С увеличением интервала реализации "m" $(T_{ck} = m t_0)$ при $S = \text{const}$ число возможных сигнальных конструкций растет (табл. 1). Информационная емкость одного кодового слова $(\log_2 N_p)$ также растет с увеличением интервала реализации "m" (табл. 3).

2. Информационная емкость одного найквистового элемента при увеличении m вначале растет (до $m \le 5$), а при $m > 5$ начинает снижаться. Для каждого "m" информационная емкость с ростом "S" увеличивается.

Как показано выше, ограничением для "S" является вероятность смещения одного момента модуляции на величину $[P(\theta = 1\Delta)] < P_3$.

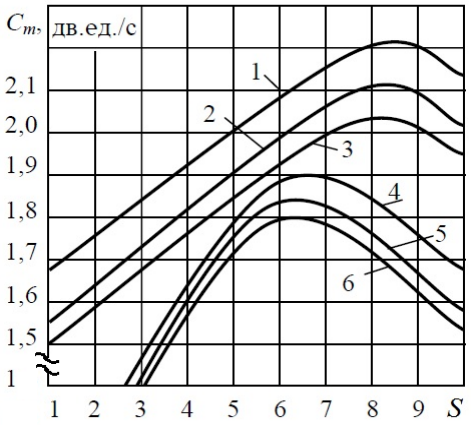

Рис. 3. Зависимости пропускной способности канала $C_i = f(S)$ при h = const, m = const

Выбор параметра "S" целесообразно ограничивать значением $S = 8$, так как при $S > 8$ увеличивается вероятность ошибочного приема одного момента модуляции даже в "хорошем" состоянии канала. Связано это с тем, что при $S > 8$ растет значение информационной емкости одного кодового слова $(\log_2 N_p)$, но еще быстрее растут потери за счет увеличения вероятности ошибки на $\theta = 1\Delta$.

Литература

1. Эффективные системы передачи информации / Н.В. Захарченко, Е.М. Рудый, А.А. Вараксин, М.А. Мамедов, М.М.Гаджиев; под ред. Н.В. Захарченко. – Баку: ЭЛМ. – 2007. – 568 с.

2. Захарченко М. В. Системи передавання даних Том 1. Завадостійке кодування / М. В. Захарченко – Одеса.: Фенікс 2009. – 447 с.

3. Захарченко Н.В. Теоретические основы оптимизации узлов и сетей / Захарченко Н.В., Мамедов М.А., Гаджиев М.М.; под ред. Н.В. Захарченко. – Баку: ЭЛМ. – 2007. – 272 с.

4. Захарченко М. В. та ін. Математичні основи оптимізації телекомунікаційних систем: підручник. За заг. ред. Захарченко М.В. - Одеса: ОНАЗ ім. О.С.Попова, 2010.– 240 с.

5. Фельдбаум А.А и др. Теоретические основы связи и управления. М.: Физматгиз – 1963. – 932 с.

6. Захарченко Н.В. Сети и системы телекоммуникаций / Н.В. Захарченко, Г.С. Гайворонская, А.И. Ещенко и др. – Киев: Техника, 2000. – Т.1. – 304 с.

References

1. Effektivnye sistemy peredachi informacii / N.V. Zaharchenko, E.M. Rudyj, A.A. Varaksin, M.A. Mamedov, M.M.Gadzhiev; pod red. N.V. Zaharchenko. – Baku: EHLM. – 2007. – 568 s.

2. Zaharchenko M. V. Sistemi peredavannya danih Tom 1. Zavadostijke koduvannya / M. V. Zaharchenko – Odesa.: Feniks 2009. – 447 s.

3. Zaharchenko N.V. Teoreticheskie osnovy optimizacii uzlov i setej / Zaharchenko N.V., Mamedov M.A., Gadzhiev M.M.; pod red. N.V. Zaharchenko. – Baku: EHLM. – 2007. – 272 s.

4. Zaharchenko M. V. ta in. Matematichni osnovi optimizaciї telekomunikacijnih sistem: pidruchnik. Za zag. red. Zaharchenko M.V. - Odesa: ONAZ im. O.S.Popova, 2010.– 240 s.

5. Fel'dbaum A.A i dr. Teoreticheskie osnovy svyazi i upravleniya. M.: Fizmatgiz – 1963. – 932 c.

6. Zaharchenko N.V. Seti i sistemy telekommunikacij / N.V. Zaharchenko, G.S. Gajvoronskaya, A.I. Eshchenko i dr. – Kiev: Tekhnika, 2000. – T.1. – 304 s.

Рецензія/Peer review : 11.2.2016 р. Надрукована/Printed : 26.3.2016 р.