УДК 681.142.003

ОПТИМИЗАЦИЯ СТРУКТУРЫ ВЫЧИСЛИТЕЛЬНОЙ СИСТЕМЫ

В.М. Захаров
(Жуковский)

Вычислительная техника уже прошла тот необходимый в каждой новой развивающейся области техники период, когда основное внимание обращается на технические параметры. Наступил период экономического проектирования [I], требующий технического решения, оптимального с экономической точки зрения, выбора оптимальных технико-экономических параметров разрабатываемой вычислительной системы в соответствии с особенностями и условиями ее эксплуатации в настоящее время и в ближайшей перспективе. Об этом свидетельствует хотя бы все увеличивающееся число публикаций, посвященных экономическим и технико-экономическим вопросам вычислительных машин [2 ÷ 9].

Так как во множестве технических параметров, характеризующих любую вычислительную систему, основным является ее производительность и из экономических параметров - себестоимость единицы полезного времени работы, то прежде всего необходимо оптимизировать соотношение между этими параметрами. В соответствии с этим должен формулироваться и критерий оптимальности вычислительной системы.

В настоящее время в качестве такого критерия применяется критерий цены эффективного быстродействия [2 ÷ 4]

$$Ц = \frac{C}{V_e \, T_M} , \qquad (I)$$

где C – приведенная сумма затрат на разработку, изготовление и эксплуатацию системы за время морального износа T_M;

v_e – эффективное быстродействие системы.

Но соотношение (I) является оценкой абсолютной эффективности вычислительной системы, а, как показывает практика, вычислительные системы с высоким быстродействием обладают большими возможностями на единицу затрачиваемой стоимости, чем системы с меньшим быстродействием. Поэтому в качестве критерия оптимальности вычислительной системы должна быть выбрана относительная оценка эффективности, так как только она характеризует возможности системы с точки зрения ее приближения к идеальной системе ("потолку" возможностей).

Относительная же оценка эффективности может быть получена только на основе анализа статистического материала по связи технических и экономических параметров существующих вычислительных систем.

Такой анализ проведен в работе автора [IO]. Рассмотренный достаточно обширный материал позволил увидеть закономерность связи быстродействия (производительности) вычислительных систем и их устройств и стоимости (себестоимости) единицы их рабочего времени, которая сформулирована в виде следующего закона:

БЫСТРОДЕЙСТВИЕ (ПРОИЗВОДИТЕЛЬНОСТЬ) ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМ И ИХ УСТРОЙСТВ ПРОПОРЦИОНАЛЬНО КВАДРАТУ СТОИМОСТИ (СЕБЕСТОИМОСТИ) ЕДИНИЦЫ ИХ РАБОЧЕГО ВРЕМЕНИ.

Существенным аргументом, подтверждающим справедливость этого закона, является то, что он выведен на основании данных, не связанных какими-либо субъективными факторами и полученных в совершенно разных технических и экономических условиях – в СССР, США, Канаде и ФРГ.

Используя этот закон, мы придем к следующей формулировке критерия оптимальности, отражающей стремление "выжать" из данной структуры вычислительной системы максимально возможное быстродействие при минимальной стоимости или, другими словами, получить быстродействие системы, соответствующее ее стоимости:

$$K_o = \frac{v_{e\,\text{реальное}}}{v_{e\,\text{ожидаемое}}} \qquad (2)$$

или с точностью до коэффициента пропорциональности

$$K_o = \frac{v_{e\,реальное}}{C^2} . \qquad (3)$$

Необходимо отметить, что введенный критерий оптимальности (3), в силу специфики определения значения эффективного быстродействия вычислительной системы, имеет статистический характер, т.е. позволяет разработать систему, наилучшую в среднем.

Одной из важнейших задач проектирования вычислительных систем является выбор оптимального соотношения быстродействий устройств. В работах [7 ÷ 9] для решения данной задачи предлагаются либо эмпирические зависимости, имеющие ограниченный характер по числу устройств оптимизируемой структуры и по типам самих структур, либо задача решается по так называемому эффекту насыщения.

Ниже предлагается метод ее решения с помощью введенного критерия оптимальности.

Считаем, что нам задана блок-схема вычислительной системы и известны зависимости между быстродействием каждого устройства системы и их стоимостью, которые, как указывалось выше, имеют вид

$$v_i = \frac{1}{t_i} = e\, c_i^2 , \qquad (4)$$

где t_i – среднее время выполнения элементарной операции i-м устройством,

c_i – стоимость i-го устройства,

e – коэффициент пропорциональности.

Значение стоимости вычислительной системы в целом C определяется как сумма стоимостей всех устройств системы, т.е.

$$C = \sum_{i=1}^{n} c_i , \qquad (5)$$

где n – число устройств системы, или с учетом (4)

$$C = \sum_{i=1}^{n} a_i t_i^{-\frac{1}{2}} , \qquad (6)$$

где $a_i = e^{-\frac{1}{2}}$.

Так как значение эффективного быстродействия v_e реальное в выражении (3) мы можем охарактеризовать временем T выполнения заданной тестпрограммы при известной ее алгоритми-

ческой сложности z [2], т.е. считать

$$v_{\text{реальное}} = \frac{z}{T}, \qquad (7)$$

то критерий оптимальности (3) можно представить в виде

$$K_o = \frac{z}{T(\sum_{i=1}^{n} a_i t_i^{-\frac{1}{2}})^2}. \qquad (8)$$

Исследование этой функции относительно быстродействия устройств t_i покажет, при каком значении $\langle t_1^0; t_2^0; \dots; t_n^0 \rangle$ стоимость на вычислительную систему затрачена наиболее рационально.

Для аналитического исследования критерия оптимальности (8) необходимо знать зависимость эффективного быстродействия вычислительной системы от быстродействия устройств или в нашей постановке задачи, зависимость $T = T(t_1; t_2; \dots; t_n)$. Этот вопрос решен в [II].

Для случая однозначного определения критического пути на ориентированном графе, описывающем процесс решения задачи вычислительной системой, в [II] показано, что такая зависимость имеет вид

$$T = \sum_{i=1}^{n} N_{\text{кр}\,i} t_i, \qquad (9)$$

где $N_{\text{кр}\,i}$ – число элементарных операций i-го устройства, лежащих на критическом пути, и предложена простая методика получения ее при моделировании вычислительной системы.

Неоднозначность определения критического пути в некоторой точке $\langle t_1^u; t_2^u; \dots; t_n^u \rangle$, с одной стороны, как бы свидетельствует о том, что для функции $T = T(t_1; t_2; \dots; t_n)$ данная точка является особой, ибо какие-то частные производные $\frac{\partial T}{\partial t}$ зависят от того, каким образом стремится к нулю Δt_i. С другой стороны, анализ зависимостей $T = T(t_i)$ для некоторых t_i из [9] показывает, что, по крайней мере для выбранных независимых переменных и исследуемой структуры вычислительной системы (достаточно характерной для современного

состояния вычислительной техники в случае [9]), эти зависимости имеют непрерывные производные, т.е. многообразие всевозможных режимов работы устройств системы "сглаживает" особые точки функции.

Поэтому с точностью до вторых производных (неоднозначность критического пути, как легко видеть, проявляется во вторых производных интересующей нас функции) функцию $T = T(t_1; t_2; \ldots; t_n)$ можно определить как

$$T_{Df} = \sum_{i=1}^{n} \frac{\sum\limits_{j=1}^{P} N_{\text{кр}ij}}{P} t_i \quad , \tag{IO}$$

где коэффициенты линейного приближения определяются как среднеарифметическое $N_{\text{кр}ij}$ по всем j критическим путям, т.е.

$$\frac{\partial T}{\partial t_i} = \frac{\sum\limits_{j=1}^{P} N_{\text{кр}ij}}{P}, \qquad j = 1,2,\ldots,P. \tag{II}$$

Практически линейная зависимость $T = T(t_1; t_2;\ldots; t_n)$ вида (IO) может быть получена при моделировании системы следующим образом: как было показано в [II], в случае неоднозначного определения критического пути какое-либо k-е устройство модели системы должно получить l-ые характеристические карточки от m других устройств, которые закончили свою работу одновременно, а работа данного k-го устройства начинается как следствие окончания работы каждого из них. (Характеристической карточкой названо слово, состоящее из n слов d_{ki}; $i = 1,2,\ldots, n$.) Включим в характеристическую карточку еще одно слово $d_{k(n+1)}$, указывающее усреднением какого количества критических путей получена данная карточка. Тогда, формируя из всех m характеристических карточек одну, каждое слово которой находится по правилу

$$d_{ki} = \frac{\sum\limits_{l=1}^{m} d_{li}\, d_{l(n+1)}}{\sum\limits_{l=1}^{m} d_{l(n+1)}} \quad ; \quad i = 1,2,\ldots,n;\ l = 1,2,\ldots m;\ i \neq k;$$

$$d_{kk} = \frac{\sum\limits_{l=1}^{m} d_{lk}\, d_{l(n+1)}}{\sum\limits_{l=1}^{m} d_{l(n+1)}} \quad ;$$

и

$$d_{k(n+1)} = \sum_{i=1}^{m} d_{i(n+1)} ,$$

где $d_{\alpha\beta}$ – обозначает β-ое слово характеристической карточки, находящейся в α-ом устройстве, в устройстве, последним окончившим работу, мы получим характеристическую карточку, содержащую в своем i-ом слове ($i = 1,2,\ldots, n$) значение

$$\frac{\sum_{j=1}^{P} N_{\kappa p i j}}{P} .$$

От такой характеристической карточки легко перейти к записи линейной зависимости вида (10).

Таким образом, предлагаемый метод требует для нахождения линейного приближения $T = T(t_1; t_2; \ldots; t_n)$ (и тем самым для нахождения аналитического выражения функции критерия оптимальности $K_0 = K_0(t_1; t_2; \ldots; t_n)$ в некоторой окрестности точки $\langle t_1^u; t_2^u; \ldots; t_n^u \rangle$ одного прогона тестпрограммы на модели вычислительной системы при $t_i = t_i^u$.

Критерий оптимальности при этом формулируется согласно (8) в виде

$$K_0 = \frac{2}{\left(\sum_{i=1}^{n} a_i t_i^{-\frac{1}{2}}\right)^2 \sum_{i=1}^{n} \frac{\sum_{j=1}^{P} N_{\kappa p i j}}{P} t_i} . \tag{12}$$

Так как аналитическое выражение (12) критерия оптимальности представляет собой приближение, справедливое только в некоторой области значений $\langle t_1; t_2; \ldots; t_n \rangle$, то естественно применить для нахождения максимума K_0 метод быстрого спуска. Известно, что движение в направлении градиента – это движение по кратчайшему, наиболее крутому, пути к экстремуму.

Для случая максимизации критерия оптимальности структуры вычислительной системы проекции вектора, определяющего направление быстрейшего подъема, должны быть пропорциональны значениям частных производных K_0 по t_i в точке исследования

$\langle t_1^u;\ t_2^u;\ldots;\ t_n^u \rangle$,т.е.

$$\Delta t_i = k\,\frac{\partial K_0}{\partial t_i}\ , \tag{I3}$$

где k - положительное число.

Или из (I2)

$$\Delta t_i = k\,\frac{2}{\left(\sum\limits_{i=1}^{n} a_i t_i^{-\frac{1}{2}}\right)^2 \sum\limits_{i=1}^{n} \frac{\sum\limits_{j=1}^{P} N_{\text{кр}ij}}{P}\, t_i} \cdot$$

$$\cdot\left[-\frac{\dfrac{\sum\limits_{j=1}^{P} N_{\text{кр}ij}}{P}}{\sum\limits_{i=1}^{n}\dfrac{\sum\limits_{j=1}^{P} N_{\text{кр}ij}}{P}\, t_i} + \frac{a_i t_i^{-3/2}}{\sum\limits_{i=1}^{n} a_i t_i^{-\frac{1}{2}}}\right]. \tag{I4}$$

Изменив значения t_i^u на Δt_i (вычисленные согласно (I4) в точке исследования $\langle t_1^u;\ t_2^u;\ \ldots;\ t_n^u \rangle$),"прогоняем" тестпрограмму на модели вычислительной системы и находим новое направление для движения к экстремуму K_0 . И так далее, пока не придем к почти стационарную область, где линейное приближение интегральной зависимости эффективного быстродействия вычислительной системы от быстродействия устройств оказывается уже недостаточным. Тогда, если желательно более точное нахождение экстремума, необходимо описывать данную зависимость полиномом более высокого порядка, чем первый.

Итерационный процесс, выполняемый по предложенной методике, должен привести нас в такую точку $O\,(t_1^0;\ t_2^0;\ldots;t_n^0)$, где критерий оптимальности K_0 будет иметь максимальное значение из всех значений K_0 в области определения аргументов t_i и,следовательно,проекции вектора скорости изменения K_0 в данной точке будут равны 0. Из (I4) видим, что это выполнится только при

$$-\frac{\dfrac{\sum\limits_{j=1}^{P} N_{\text{кр}ij}}{P}}{\sum\limits_{i=1}^{n}\dfrac{\sum\limits_{j=1}^{P} N_{\text{кр}ij}}{P}\, t_i^0} + \frac{a_i t_i^{0\ -3/2}}{\sum\limits_{i=1}^{n} a_i t_i^{-\frac{1}{2}}} = 0\ . \tag{I5}$$

Проанализируем данное выражение.

Замечаем, что согласно (4), (5) и (10), (15) можно записать

$$- \frac{\frac{\sum\limits_{j=1}^{P} N_{кpij}}{P}}{T^0} + \frac{\frac{c_i^0}{t_i^0}}{C^0} = 0 , \qquad (16)$$

$$\frac{\frac{c_i^0}{t_i^0}}{C^0} = \frac{\frac{\sum\limits_{j=1}^{P} N_{кpij}}{P}}{T^0} , \qquad (17)$$

$$\frac{c_i^0}{C^0} = \frac{t_i^0 \frac{\sum\limits_{j=1}^{n} N_{кpij}}{P}}{T^0} . \qquad (18)$$

В случае однозначного определения критического пути на графе, описывающем процесс решения задачи на вычислительной системе, т.е. $P = 1$, равенство (18) принимает вид

$$\frac{c_i^0}{C^0} = \frac{N_{кpi} t_i^0}{T^0} . \qquad (19)$$

А это означает, что доля стоимости устройства в общей стоимости вычислительной системы равна доле времени работы данного устройства на критическом пути во всем времени работы системы.

Для выяснения значения (18) для общего случая рассмотрим частный случай процесса (рис. 1) и временную диаграмму Джан-

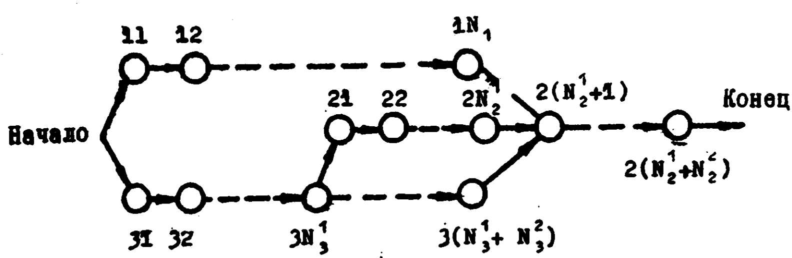

Рис. 1

Через ij обозначено - i-ое устройство работает j-ый раз

та, соответствующую ему (рис. 2). Для данного процесса соглас-

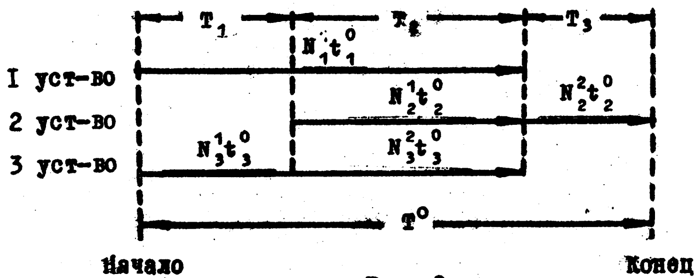

Рис. 2

но (I8) получаем

$$\frac{c_1^0}{C^0} = \frac{\frac{1}{3} N_1 t_1^0}{T^0} , \tag{20}$$

$$\frac{c_2^0}{C^0} = \frac{(\frac{1}{3} N_2^1 + N_2^2) t_2^0}{T^0} , \tag{2I}$$

$$\frac{c_3^0}{C^0} = \frac{(\frac{2}{3} N_3^1 + \frac{1}{3} N_3^2) \; t_3^0}{T^0} . \tag{22}$$

Перепишем (20) ÷ (22) в следующем виде:

$$\frac{c_1' + c_1'' + c_1'''}{C^0} = \frac{\frac{1}{3} N_1' \, t_1^0}{T^0} + \frac{\frac{1}{3} N_1'' \, t_1^0}{T^0} + \frac{0}{T^0} , \tag{23}$$

$$\frac{c_2' + c_2'' + c_2'''}{C^0} = \frac{0}{T^0} + \frac{\frac{1}{3} N_2^1 t_2^0}{T^0} + \frac{N_2^2 t_2^0}{T^0} , \tag{24}$$

$$\frac{c_3' + c_3'' + c_3'''}{C^0} = \frac{\frac{2}{3} N_3^1 t_3^0}{T^0} + \frac{\frac{1}{3} N_3^2 t_3^0}{T^0} + \frac{0}{T^0} , \tag{25}$$

где N_1' и N_1'' – число операций первого устройства соответственно в интервале T_1 и T_2,

$$\begin{aligned} c_1^0 &= c_1' + c_1'' + c_1''' , \\ c_2^0 &= c_2' + c_2'' + c_2''' , \\ c_3^0 &= c_3' + c_3'' + c_3''' . \end{aligned} \tag{26}$$

Потребуем, чтобы

$$
\begin{aligned}
&\frac{c'_1}{C^0}=\frac{\frac{1}{3}N'_1t^0_1}{T^0}; &&\frac{c''_1}{C^0}=\frac{\frac{1}{3}N'_1t^0_1}{T^0}; &&\frac{c'''_1}{C^0}=\frac{0}{T^0};\\
&\frac{c'_2}{C^0}=\frac{0}{T^0}; &&\frac{c''_2}{C^0}=\frac{\frac{1}{3}N'_2t^0_2}{T^0}; &&\frac{c'''_2}{C^0}=\frac{N^2_3t^0_2}{T^0}; \qquad (27)\\
&\frac{c'_3}{C^0}=\frac{\frac{2}{3}N_3t^0_3}{T^0}; &&\frac{c''_3}{C^0}=\frac{\frac{1}{3}N^2_3t^0_3}{T^0}; &&\frac{c'''_3}{C^0}=\frac{0}{T^0}.
\end{aligned}
$$

Легко видеть, что такое разбиение стоимости каждого устройства вполне допустимо.

Но тогда

$$
\left.
\begin{array}{l|l|l}
\dfrac{c'_1}{C^0}=\dfrac{\frac{1}{3}T_1}{T^0} & \dfrac{c''_1}{C^0}=\dfrac{\frac{1}{3}T_2}{T^0} & \dfrac{c'''_1}{C^0}=0\\[3ex]
\dfrac{c'_2}{C^0}=0 & \dfrac{c''_2}{C^0}=\dfrac{\frac{1}{3}T_2}{T^0} & \dfrac{c'''_2}{C^0}=\dfrac{T_3}{T^0}\\[3ex]
\dfrac{c'_3}{C^0}=\dfrac{\frac{2}{3}T_1}{T^0} & \dfrac{c''_3}{C^0}=\dfrac{\frac{1}{3}T_2}{T^0} & \dfrac{c'''_3}{C^0}=0
\end{array}
\right\} \qquad (28)
$$

Просуммируем равенства, находящиеся в одной колонке (28):

$$
\frac{c'_1+c'_2+c'_3}{C^0}=\frac{T_1}{T^0}; \quad \frac{c''_1+c''_2+c''_3}{C^0}=\frac{T_2}{T^0}; \quad \frac{c'''_1+c'''_2+c'''_3}{C^0}=\frac{T_3}{T^0} \qquad (29)
$$

или

$$
\frac{c'_1+c'_2+c'_3}{T_1}=\frac{C^0}{T^0}; \quad \frac{c''_1+c''_2+c''_3}{T_2}=\frac{C^0}{T^0}; \quad \frac{c'''_1+c'''_2+c'''_3}{T_3}=\frac{C^0}{T^0}. \qquad (30)
$$

Откуда можно сделать вывод, что в оптимизированной структуре стоимость вычислительной системы C^0 равномерно распределена по времени работы системы T^0, т.е. любые равные моменты работы системы имеют одну и ту же стоимость и эта стоимость рас-

пределяется между устройствами так, чтобы привести к максимально возможной параллельной работе устройств.

Но это согласуется с интуитивным представлением об оптимальной структуре вычислительной системы, что позволяет считать, что выбор критерия оптимальности правилен.

Более же строгая оценка предлагаемой методики может быть сделана только по результатам ее применения на практике.

Литература

I. Д.С. Львов. Основы экономического проектирования машин. М., Изд-во "Экономика", 1966.

2. В.М. Глушков. Два универсальных критерия эффективности вычислительных машин. ДАН УССР, 1960, № 4.

3. А.В. Шилейко. Методика выбора оптимальной структуры цифровой модели. - Автоматика и телемеханика, 1961, т. ХХП, № I.

4. А.Ш. Чадунели. Об определении эффективности модернизации специализированных математических машин. В сб. трудов семинара по математическим методам моделирования и теории электрических цепей. Киев, Изд-во "Наукова думка", 1964, вып. 2.

5. H.R.I.Grosch. The Future of computing. - Indus. and Eng. Chemistry, 1958, vol.50, N 11,

6. H.Kittel, P.Mertens. Einige quantitative Untersuchungen zűr Grössendegression von Datenverarbeitungsanlagen. - Electron.Datenverarb., 1965, Bd.7, N 6.

7. Е.И. Мамонов. Вопросы расчета быстродействия устройств и оценка производительности ЦВМ. - Вычислительная техника, МИФИ, 1961, вып. 2.

8. Я.А. Хетагуров и др. Об оценке эффективности устройств ЦВМ. - Изв. ВУЗов, Приборостроение, 1965, № 6.

9. Проектирование сверхбыстродействующих систем "СТРЕТЧ" (под ред. Б.Бухгольца). М., Изд-во "Мир", 1965.

10. В.М. Захаров. Взаимосвязь быстродействия ЦВМ и ее устройств и их стоимости. Готовится к печати.

II. В.М. Захаров, А.Н. Стихов. Об интегральной зависимости скорости ЦВМ от быстродействия устройств. — Тезисы доклада на ХХШ Всесоюзной научной сессии НТОРЭ им. А.С. Попова, М., 1967.