ワークファイルの管理方式とそのモデルの解析

深川幸紀　　積山洋子　　矢嶋虎夫
（西日本工大）　（九工大）　（九工大）

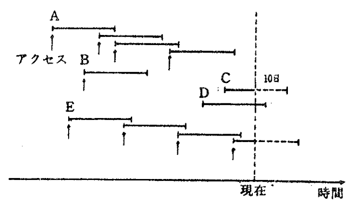
1. まえがき

本文では，情報処理教育センターや計算センター等におけるワークファイル（特に演習用プログラムファイル）の管理方式とこのシステムのモデルについて検討した。ユーザーが計算機システムを使用する時ワークファイルを用いることが多い。又複数台の端末を用いてオンラインでプログラム教育を行なう場合，教育効果は大きい。この様な場合，ユーザー毎にワークファイルあるいはプログラムファイルが作成されるが，全ユーザー（潜在的ユーザーも含める）のためのファイル（エリア）を準備する場合大きな容量のファイル（多数のファイル）が必要となる。例えばプログラムファイルの場合，1000人の学生が1人あたり100枚のプログラムカードあるいはデータカードを記憶するとすれば，約80メガバイトの容量が必要となる。このような場合，中小規模システムでは負担が大きい。ここで，ある程度小さい容量のファイルエリアを用いて，ユーザーにあまり不便をかけずにシステム運用をしてゆくための簡単な方法を提案し，この方式をある待ち行列でモデル化出来ることを示し，これを用いてファイルの使用中の容量（ファイル数）に関する確率や平均値を求めた。

2. 管理方式とモデル

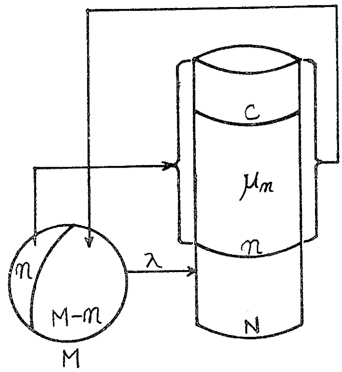
2.1 方式とモデル

ファイルの管理運用方式は，ある1人のユーザーのファイル（あるいはファイル区画・・・以後ファイルとファイル区画を区別しない）を作成した時点からファイル有効期間を固定し，もしこの間にファイルへのアクセスがあればこの時点から有効期間を決められた固定の長さ延長する方式である。計算機システムは個々のファイルの最新アクセス日付と今日の日付の差を計算し，一定期間（有効期間と同じ長さ）をこしていれば，このファイルをクリアして他のユーザーのために空きエリアとする。すなわち過去一定期間アクセス（使用）されていないファイルを消す方法で，簡単であり，かつ自然な方法である。ユーザー側から考えると，ファイルをアクセスし続ければ，消されることはないシステムである。さて，このような運用方式をとる時，各個のユーザーファイルの使用期間は図－1に示すようにばらつく。演習用プログラムファイルについていえば，問題の難易によるファイル有効期間の重みづけや持ち点制（CPU時間，LP出力枚数やファイル使用期間等）による演習方式の場合はさらにばらつくであろう。又図－1のユーザーEのように長期間ファイルを使用してきたが，現在から将来まであとどのくらい使用し続けるかは，ユーザーEのアクセス状況によっていてわからない。このような様相はファイル使用期間の分布を指数分布と仮定してもよいと考えられる。



図－1　ファイル使用期間

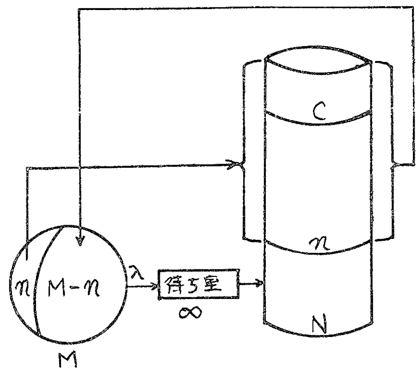
ここでファイルの運用方式を，個人毎のファイルをサーバーとみなし，ファイル使用期間が，指数分布をもつサービス時間であるような待ち行列でモデル化を行なう。そして使用中のファイル数が大きくなれば固定された有効期間を短かくするように制御をする（今日の日付と最新ファイルアクセス日付の差を短かくしてファイルを消去する）時，この制御が処理率の平均値を変えると考えられ，ファイル運用管理方式を，有限呼源複数サーバーで平均処理率が動作中のサーバー数に比例するような待ち行列で近似した。このモデルを図－２と図－３に示す。図－３の待ち室の意味は，ファイルが全て使用中の時，ファイルを新規に使いたいユーザーは毎日システムにファイル要求をするモデルであると考えられる。ここではファイル数がある固定のC以下の時は消去動作を行なわないモデルを取り上げる。



図－２　有限呼源モデルA

２．２　有限呼源モデルAの平衡方程式

全ユーザー数をM，ファイル数をNとし，ユーザーのファイル要求率は平均をλとするポアソン分布とし，ファイル使用時間は平均をμ_n^{-1}とする指数時間分布とする。ファイル数がC以下の時消去をしない。平衡状態における使用中のファイル数がnである確率をP_nとする。



図－３　有限呼源モデルB

２．２．１　平衡方程式

$$0 = -M\lambda P_0 \quad \text{---} \ (1)$$

$$0 = -(M-n)\lambda P_n + \{M-(n-1)\}P_{n-1} \quad (1 \leqq n \leqq C-1) \quad \text{---} \ (2)$$

$$0 = -(M-C)\lambda P_C + \{M-(C-1)\}\lambda P_{C-1} + (C+1)\mu_{C+1}P_{C+1} \quad \text{---} \ (3)$$

$$0 = -\{(M-n)\lambda + n\mu_n\}P_n + \{M-(n-1)\}\lambda P_{n-1} + (n+1)\mu_{n+1}P_{n+1} \quad \text{---} \ (4)$$

$$C+1 \leqq n \leqq N-1$$

$$0 = -\{(M-N)\lambda + N\mu_n\}P_N + \{M-(N-1)\}\lambda P_{N-1} \quad \text{---} \ (5)$$

(1)，(2)式より $P_0 = P_1 = \text{---} = P_{C-1} = 0$．（３）式より $P_{C+1} = \frac{M-C}{C+1}\rho_{C+1}P_C$．　（６）
ここに $\rho_i = \lambda/\mu_i$．（４）（６）式を用いて漸下的に

$$P_n = \frac{\binom{M}{n}}{\binom{M}{C}}\prod_{i=C+1}^{n}\rho_i P_C \quad (C+1 \leqq n \leqq N) \quad \text{---} \ (7)$$

$$\sum_{n=C+1}^{N} P_n + P_C = 1 \text{ から } P_C = \left\{ 1 + \binom{M}{C}^{-1} \sum_{n=C+1}^{N} \binom{M}{n} \prod_{i=C+1}^{n} \rho_i \right\}^{-1} \qquad (8)$$

平均値は $\sum_{n=C}^{N} nP_n$ から計算出来るが簡単な形では表わせない。

2.2.2 有限呼源モデルBの平衡方程式

このモデルでは，待ち室人数は制御に関係せず，ファイル数のみが関係する。計算機は待ち人数を知ることが出来ないからである．

$$0 = -M\lambda P_0 \qquad (9)$$

$$0 = -(M-n)\lambda P_n + \{M-(n-1)\}\lambda P_{n-1} \qquad (10)$$
$$(1 \leqq n \leqq C-1)$$

$$0 = -(M-C)\lambda P_C + \{M-(C-1)\}\lambda P_{C-1} + (C+1)\mu_{C+1} P_{C+1} \qquad (11)$$

$$0 = -\{(M-n)\lambda + n\mu_n\} P_n + \{M-(n-1)\}\lambda P_{n-1} + (n+1)\mu_{n+1} P_{n+1} \qquad (12)$$
$$(C+1 \leqq n \leqq N-1)$$

$$0 = -\{(M-n)\lambda + N\mu_N\} P_n + \{M-(n-1)\}\lambda P_{n-1} + N\mu_N P_{n+1} \qquad (13)$$

$$0 = -N\mu_N P_M + \lambda P_{M-1} \qquad (14)$$

(9)，(10)式より $P_0 = P_1 = \cdots = P_{C-1} = 0$．残りの式より漸下的に

$$P_n = \frac{\binom{M}{n}}{\binom{M}{C}} \prod_{i=C+1}^{n} \rho_i P_C \quad (C+1 \leqq n \leqq N) \qquad (15)$$

$$P_n = N^{N-n} \frac{n!}{N!} \frac{\binom{M}{n}}{\binom{M}{C}} \frac{\prod_{i=C+1}^{N} \rho_i}{\rho_N^{N-n}} P_C \quad (N \leqq n \leqq M) \qquad (16)$$

P_Cは正規化条件 $P_C + \sum_{n=C+1}^{M} P_n = 1$ より求まる．

$$P_C = \left\{ 1 + \sum_{n=C+1}^{N} \frac{\binom{M}{n}}{\binom{M}{C}} \prod_{i=C+1}^{n} \rho_i + \sum_{n=N+1}^{M} N^{N-n} \frac{n!}{N!} \frac{\binom{M}{n}}{\binom{M}{C}} \frac{\prod_{i=C+1}^{N} \rho_i}{\rho_N^{N-n}} \right\}^{-1} \qquad (17)$$

2.2.3 無限呼源行列（待ち室なし）モデルについて

このモデルについては，平衡方程式は略すが漸下的にP_nを求めることが出来る。ここでは結果だけを示す。

$$P_n = \frac{C!}{n!} \prod_{i=C+1}^{n} \rho_i P_C \quad (C+1 \leqq n \leqq N) \qquad (18) \qquad P_C = \left\{ 1 + \sum_{n=C+1}^{N} \frac{C!}{n!} \prod_{i=C+1}^{n} \rho_i \right\}^{-1}$$

2．2．4　無限呼源（待ち室付き）行列モデル

$$P_n = \frac{C!}{n!}\prod_{i=C+1}^{n}\rho_i P_C \qquad C+1 \leqq n \leqq N \qquad --- (20)$$

$$P_n = \frac{\rho_N^{n-N}}{N^{n-N}}\frac{C!}{N!}\prod_{i=C+1}^{N}\rho_i P_C \qquad N \leqq n \qquad --- (21)$$

P_Cは正規化条件から求められる。

$$P_C = \left\{1 + \sum_{n=C+1}^{N-1}\frac{C!}{n!}\prod_{i=C+1}^{n}\rho_i + \sum_{n=N}^{\infty}\frac{\rho^{n-N}}{N^{n-N}}\frac{C!}{N!}\prod_{i=C+1}^{N}\rho_i\right\}^{-1} \quad --- (22)$$

さてここで有限呼源モデルで$M\to\infty$，$\lambda\to 0$かつ$M\lambda\to\lambda'$とすると式（15）が式（18）になることを示す。（15）式のP_nについて極限をとると

$$\lim_{\substack{M\to\infty\\ \lambda\to 0\\ M\lambda\to\lambda'}} P_n = \lim_{\substack{M\to\infty\\ \lambda\to 0\\ M\lambda\to\lambda'}} \frac{\binom{M}{n}}{\binom{M}{C}}\prod_{i=C+1}^{n}\rho_i P_C = \lim_{\substack{M\to\infty\\ \lambda\to 0\\ M\lambda\to\lambda'}} \underbrace{(M-C)\cdot(M-C-1)\cdots(M-n+1)}_{n-C\text{個}}\lambda^{n-C}\frac{C!}{n!}\prod_{i=C+1}^{n}\mu_i^{-1}P_C$$

$$= \lim_{\substack{M\to\infty\\ \lambda\to 0\\ M\lambda\to\lambda'}}(M\lambda-\lambda C)(M\lambda-\lambda C-\lambda)\cdots(M\lambda-n\lambda+\lambda)\frac{C!}{n!}\prod_{i=C+1}^{n}(\mu_i^{-1})P_C = \lambda'^{n-C}\frac{C!}{n!}\prod_{i=C+1}^{n}(\mu_i)^{-1}P_C = (18)$$

他の式についても有限行列の式は無限呼源行列の式に極限移行操作より等しくなる。

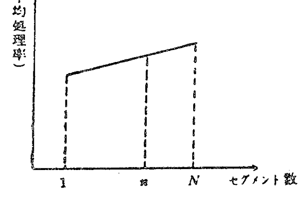
3．数値計算

ここではP_nを用いて，主として使用中のファイル数の平均値を計算し，処理率とファイル数nの関係を変えて，この効果をしらべた。この関係は直線制御A，Cと階段制御である。以下のセグメントの意味はファイル数（区画数）と同じ。

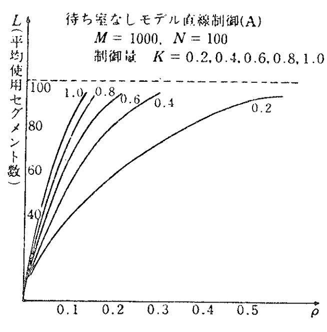
3．1　直線制御A（固定セグメント $C=0$）

$$\mu_n = \{(\mu_N-\mu_1)n + (N\mu_1-\mu_N)\}/(N-1) = \{(1-K)n + (NK-1)\}/(N-1)$$

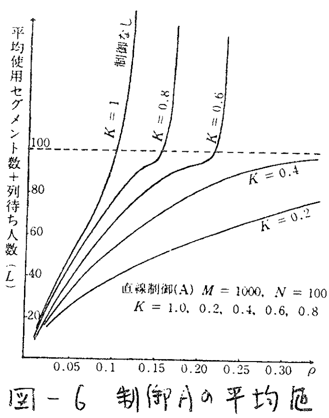
で　$K = \dfrac{\mu_1}{\mu_N}$　（$K \leqq 1$）。



図－4　制御A



図－5　制御Aの平均値



図－6　制御Aの平均値

3．2　直線制御C（固定セグメントC＝0）

$$\prod_{i=1}^{n}\mu_n=\mu_1^{n}(n-K)!\quad(n>K),\quad \prod_{i=1}^{n}\mu_n=\mu_1^{n}\quad(1\leqq n\leqq K)$$

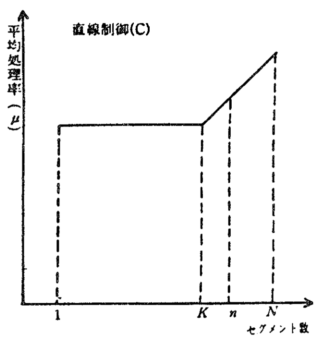
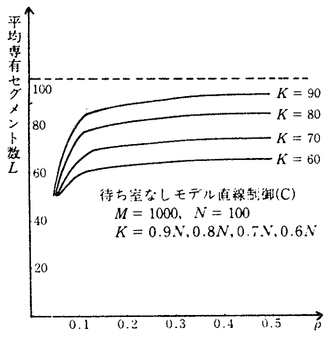
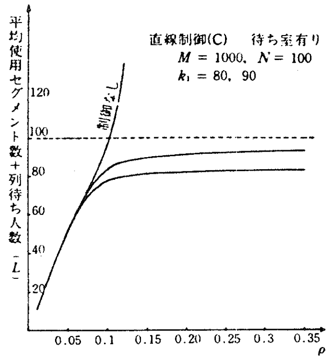


図-7　制御C



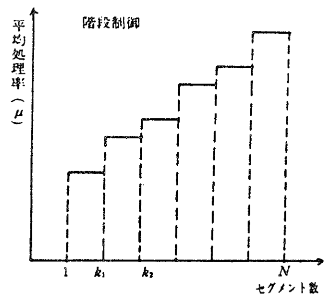
図－8　待ち室なし行列平均値



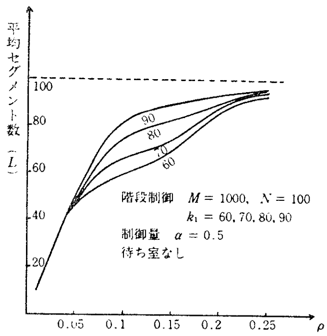
図－9　待ち室付き行列平均値

3．3　階段制御（固定セグメントC＝0）

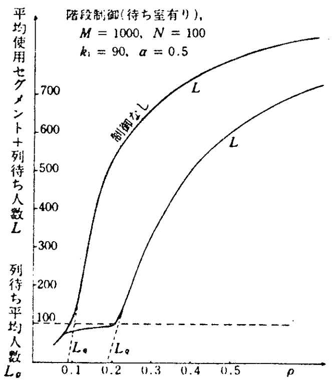
$\alpha_i=\dfrac{\mu_1}{\mu_i}$　（$\alpha_i\leqq 1$）でα_iが小さいほど制御が強い。



図－10　階段制御

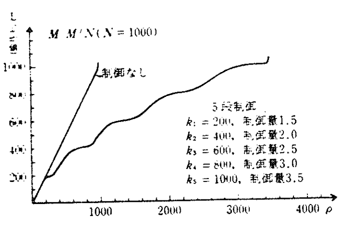


図－11　待ち室なし行列平均値

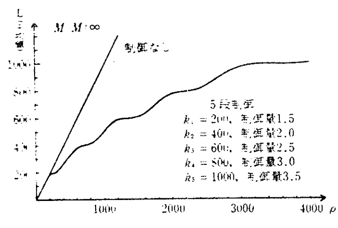


図－12　待ち室付き行列平均値

3．3　階段制御（無限呼源モデル）（固定セグメント　C＝0）



図－13　無限呼源(M/M/S)平均値



図－14　無限呼源(M/M/∞)平均値

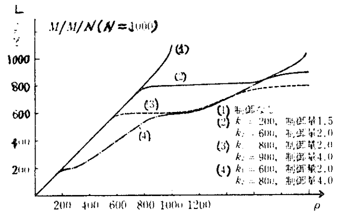


図-15 無限呼源(M/M/S)平均値

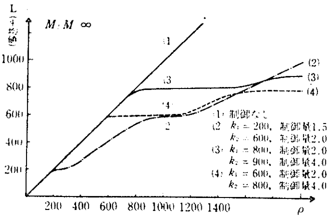


図-16 無限呼源(M/M/∞)平均値

3．5　階段制御（有限呼源モデル）
固定セグメント　C＝20，40　M＝1000

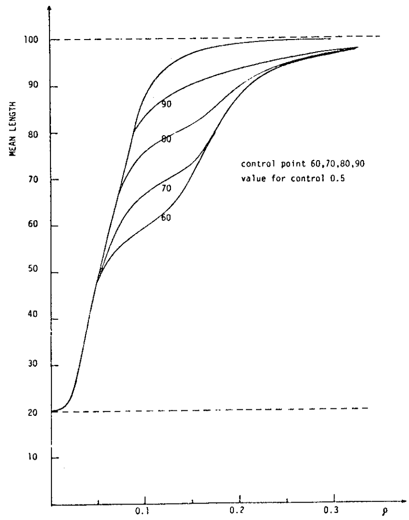


図-17　待ち室なしモデル平均値
（C＝20）

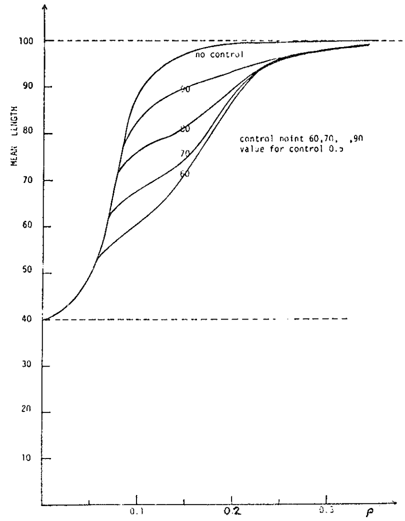


図-18　待ち室付きモデル平均値
（C＝40）

3．6　プログラムファイルの推移

図-19は九工大情報処理教育センターのファイル数の推移である。ファイルは690人分作成されている。1ケ月の有効期間を設定していて，ファイルをクリアするのは5で割り切れる日のみ行なっている。ファイル数れで処理率を制御していない。

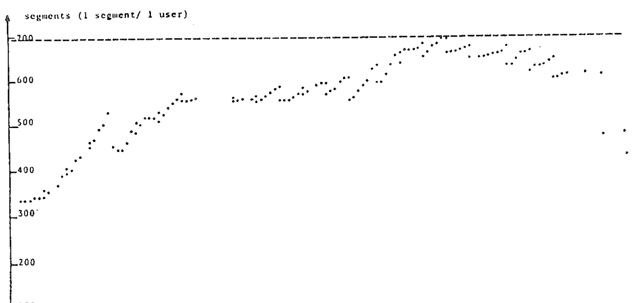


図-19 演習用プログラムファイル推移

4. むすび

簡単なワークファイル管理方式を提案し，この方式ではファイル使用期間分布が指数分布で近似出来ることを用いて，M／M／S待ち行列でモデル化を行なった。そして使用中のファイル数でファイル処理率（ファイル使用期間）を変化させる時，平衡状態の確率が得られ，これから平均値を計算した。この結果処理率の制御は1段階の階段制御で充分であることが推定された。ここでのべた管理方式は，九工大情報処理教育センターで，演習用プログラムファイルを対象として適用されている。本方式は簡単で，省力化にもなっていて，システム全体の運用は支障がないと思われる。1人のユーザーが使用するファイル容量が可変の時や到着率が時間によって変動するような場合の検討を行なう必要があろう。

＜参考文献＞

（1）深川他：“集団情報処理教育におけるプログラム用ファイル容量設定に関する一考察”，九工大研究報告（工学），NO.36，昭和53年.
（2）深川他：“情報処理教育用プログラムファイルの容量制御について”，電気四学会九支連，昭和53年
（3）Saaty：“Element of Queuing Theory”，McGRAW-HILL，1961
（4）本間：“待ち行列の理論”，理工学社，1966
（5）鈴本：“待ち行列，裳華房，1972
（6）寺本他訳：“生物学における確率過程の理論”，産業図書，1978

＜参　考＞

本文における数値計算に用いられた組み合わせ（$\binom{m}{n}$）（Combination）の計算を行なう副プログラムを示す。本文での計算のなかに１０００！等の巨大な数値が出現し，そのために筆者の１人によってこの副プログラムが作成された。

このサブルーチンの方法を（$\binom{10}{3}$）の例で説明する。

$$\binom{10}{3} = \frac{10!}{3!7!} = \frac{10\times 9\times 8}{3\times 2\times 1}$$

であり，これの１０を底とする対数をとる。

$$\begin{aligned}\log_{10}\binom{10}{3} &= \log_{10}10 - \log_{10}3\\ &\quad + \log_{10}9 - \log_{10}2\\ &\quad + \log_{10}8 - \log_{10}1\\ &= yy.fff\cdots f\end{aligned}$$

$SUM = yy.fff\cdots f$

$IA = IFIX(SUM) = yy$

IA は 10^{yy} を表わす.

$B = .fff\cdots f$

$A = 10.0^{B} = 10.0^{.fff\cdots f}$

$\therefore\ 1.0 \leqq A < 10.0$

（$\binom{1000}{500}$）等の計算結果を右に示す。本サブルーチンでyyはFORTRANのの整数の　囲を取り扱かっているので大きな数の階乗や組み合わせの計算も可能であるが精度はこの場合単精度である。倍精度の副プログラムも簡単に作成出来る。

```
      SUBROUTINE CMB(M,N,A,IA)
C
      IF(N.EQ.0) GO TO 1
      MH=M/2
      NA=N
      IF(NA.GT.MH) NA=M-NA
      SUM=0.0
      DO 10 I=1,NA
   10 SUM=SUM+ALOG10(FLOAT(M-NA+I))-ALOG10(FLOAT(I))
      IA=IFIX(SUM)
      B=SUM-FLOAT(IA)
      A=10.0**B
      RETURN
    1 IA=0
      A=1.0
      RETURN
      END
```

```
      M=1000
      DO 1 N=10,500,10
      CALL CMB(M,N,A,IA)
      WRITE(6,200) M,N,A,IA
  200 FORMAT(1H ,' M =',I6,5X,'N=',I6,F15.7,'E',I6)
    1 CONTINUE
      STOP
      END
```

```
 M =   1000     N=    10     2.6338634E    23
 M =   1000     N=    20     3.3931971E    41
 M =   1000     N=    30     2.4276876E    57
 M =   1000     N=    40     5.5539045E    71
 M =   1000     N=    50     9.4470463E    84
 M =   1000     N=    60     1.9709597E    97
 M =   1000     N=    70     7.0260334E   108
 M =   1000     N=    80     5.4203091E   119
 M =   1000     N=    90     1.0794516E   130
 M =   1000     N=   100     6.3659544E   139
 |   |  |  |  |  |  |  |  |  |  |  |  |  |  |
 M =   1000     N=   400     4.7002306E   290
 M =   1000     N=   410     2.1857452E   292
 M =   1000     N=   420     6.7317038E   293
 M =   1000     N=   430     1.3761692E   295
 M =   1000     N=   440     1.8695135E   296
 M =   1000     N=   450     1.6934004E   297
 M =   1000     N=   460     1.0233164E   298
 M =   1000     N=   470     4.1464434E   298
 M =   1000     N=   480     1.1215143E   299
 M =   1000     N=   490     2.0385723E   299
 M =   1000     N=   500     2.4846411E   299
```