選択問題の面積時間複雑度

5X-7

安浦　寛人　　　Clark D. Thompson
（京都大学工学部）　（Univ. of California, Berkeley）

1．まえがき

選択問題は，n個の整数の集合の中でℓ番目に小さな要素を見つける問題である．逐次計算モデルの上では，ソーティングに較べ選択問題は真に少ない計算量で解けることが知られている．ここでは，VLSIモデル上での選択問題の面積時間複雑度について考え，並列計算においても選択問題がソーティングよりも少ない計算量で解けることを示す．

2．準備

ここでは配線層の数が制限されたVLSIモデルを考える[1]．また，データの各ビットの入出力される位置および時刻は固定されており（When/Where-oblivious），入力される各ビットは１回しか印加されない（Semelective）と仮定する．問題の複雑さは，問題を解くVLSI回路の面積Aと計算時間Tの積ＡＴで評価する．

ℓ選択問題を，n個のkビット２進整数の集合の中からℓ番目に小さな要素を見つける問題と定義する．ここに，n，k，ℓは固定であり，上述の入出力データに対する制限以外，入出力データの形式は自由とする．$\ell=1$の場合が最小選択問題であり，$\ell=n/2$の時が中央値選択問題である．また，ℓを固定せず動的に指定できる問題を，動的選択問題と呼ぶ．

3．選択問題の面積時間積の下界

ℓ選択問題の面積時間積の下界としては，入力のバンド幅から$AT \geqq nk$が容易に得られる．さらに，出力の最上位ビットはすべての入力に依存することから，$AT^{\alpha}=\Omega(nk(\log nk)^{\alpha-1})$（$\alpha \geqq 1$）が得られる．動的選択問題は明らかに$\ell$選択問題より難しいので，これらは同時に動的選択問題の下界でもある．また，$k=1$の場合，ℓ選択問題は対称論理関数の計算に帰着できるので，面積時間積の下限は$AT^{\alpha}=\Theta(n(\log n)^{\alpha-1})$であることが導ける[2]．次節では，ある範囲の$\ell$や$k$に対しては，$AT=\Theta(nk)$が選択問題の面積時間積の下限となることを示す．

4．選択問題の面積時間積の上界

1）ラディックス選択法　各入力データの最上位ビットを並列に読み込み，その中の０の数を並列計数回路でカウントする．０がℓ個以上あれば，解の最上位ビットも０であるから，０を出力する．最上位が１であったデータは以後の計算対象から除外する．０がℓ個未満であれば，１を出力し，最上位が０のデータは以後の計算対象から除外する．また，ℓから０の数を引いておく．次に，第２桁を読み込んで，計算対象となっているデータについて同様の処理を行なう．各ステップ毎に，入力データが１桁ずつ読み込まれ，解の１桁が出力される．回路は，n入力の並列計数回路と$\log n$ビットの比較器，減算器から構成される．Ｈ木のレイアウトを採用すれば，面積は$O(n)$となる．計算時間は，１ステップが$O(\log n)$かかるから，$O(k\log n)$となる．よって，$AT=O(nk\log n)$となる．この方法は，入力を$O(\log\log n)$ビットまとめて取り扱うことにより，パイプライン化することができ，計算時間を$O(k\log n/\log\log n)$まで改善できる．しかし，面積は$\log\log n$倍になるため，ＡＴ積は改善されない．この回路は，ℓに無関係に構成でき，動的選択回路としても利用できる．

2）最小選択回路　最小選択問題（$\ell=1$）に対しては，$AT=O(nk)$の回路が作れる．回路は，$nk/(k+\log n)$個の葉節点を入力ポートとし，各内部節点が直列比較器である完全２分木として構成される．さらに，根節点には，kビットのシフトレジスタと直列比較器が接続される．入力データは，$nk/(k+\log n)$個ずつを処理単位として，最上位ビットから入力され，パイプライン処理によって各処理単位の最小要素が根節点から出力される．これと，シフトレジスタに保存されている

Area-Time Complexity of Selection Problem
Hiroto YASUURA[1], Clark D. THOMPSON[2]
1 Kyoto University, 2 University of California

以前の最小値を較べ，小さい方をシフトレジスタに残す．処理は完全にパイプライン化できるので，O（k＋log$n$）の時間で最小要素が見つかる．面積は，H木のレイアウトを採用することによって，O（nk／（k＋log$n$））とできるので，面積時間積は下界と一致しO（nk）となる．この回路は，葉節点にℓ個のデータに対するソータ，直列比較器の替わりにマージ回路を用いることにより，ℓ選択回路に一般化できる．面積はO（$nk\ell$／（k＋logℓ log$n$）），計算時間はO（k＋logℓ log$n$）となる．AT＝O（$nk\ell$）であるので，ℓ＝o（log$n$）の場合，1）よりも面積時間積が小さくなる．

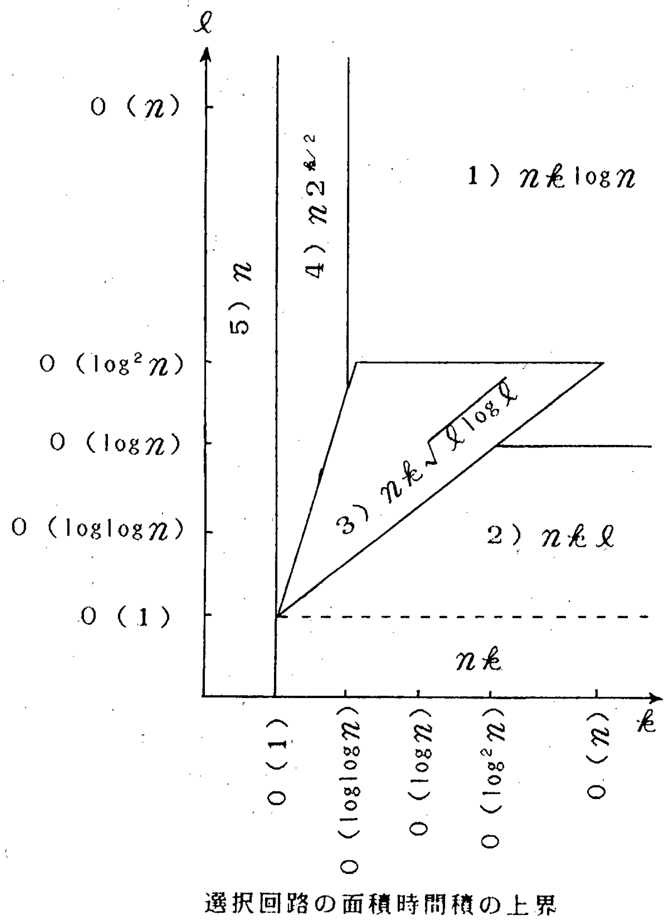
3）バブルソータによる選択回路　ℓ個のセルよりなるバブルソータはℓ選択回路として利用できる．面積はO（$k\ell$），時間はO（n＋k）となる．さらに，ColeとSiegelのAT最適ソータ[3]を用いると，面積はO（$k\ell$）ままで，時間をO（$n\sqrt{\log \ell}/\sqrt{\ell}$）まで改善できる．AT＝O（$nk\sqrt{\ell \log \ell}$）となるが，$\log\log \ell < \log k < \sqrt{\ell \log \ell} - \log \ell$の条件が付く．この条件の範囲で，かつ$\ell$が小さい時，1）や2）よりも面積時間積は小さくなる．

4）面積最小選択回路　Durisらの面積最小選択回路[4]を用いると，面積はA＝O（min｛2^k，ℓ｝（｜k－logℓ｜＋1）），時間はT＝O（nk）となる．さらに，ColeとSiegelのデータ圧縮[3]の手法を組み合わせると，k＜log2ℓの範囲で，A＝O（2^k（log4ℓ－k）），T＝O（$n2^{-k/2}$／（log4ℓ－k））とできる．面積時間積は，O（$n2^{k/2}$）となり，kが小さな領域では，上記のいずれの回路よりも小さくなる．

図に，本節の結果をまとめる．kやℓが大きなところでは，1）の回路が最良の上界を与える．2）の回路は，ℓが小さなところで有効であり，特に，ℓ＝O（1）のときは，下界と一致する．3）の回路はkとℓが共に小さな部分に有効であり，4）はkが小さな所で最良となる．

5．あとがき

図の1）～3）の部分では，ℓ選択問題の上界がColeとSiegelの与えたソーティングの下界[3]より小さくなっている．よって，この部分ではℓ選択問題がソーティングより真に複雑さの低い問題となっていることがわかる．また，5）の領域ではℓ選択問題とソーティングは同じ複雑さを持つ．よって4）の領域のどこかに，ℓ選択問題がソーティングより簡単になる境界があると思われる．最小選択問題に近い領域（ℓ＝O（1））では，2）の回路が最適であり，面積時間積の下限を達成している．また，5）の領域でも上界と下界が一致している．しかし，その他の領域ではまだ上界と下界の間にギャップがある．動的選択問題に対しては，1）の回路により，AT＝O（nklog$n$）である．



選択回路の面積時間積の上界

参考文献

1. Thompson, Ph.D Dissertation, CMU-CS-80-140, 1980
2. Wada et.al., IEEE Trans. on Comput. C-33, 5, 1984
3. Cole and Siegel 26th FOCS, 1985
4. Duris et.al., to appear, Algorithmica, 1986