Abstract

近年，数値シミュレーションの複雑化および大規模化に伴い，数百万行を超える規模の係数行列の連立一次方程式を高速に求解することが求められている．有限要素法などで離散化された偏微分方程式やグラフ解析は，行列の要素の多くが 0 である疎行列を係数行列とした連立一次方程式として表わされる．その求解には直接解法と反復解法が用いられる．しかし，直接解法はその計算量の多さや疎行列に対して完全 LU 分解をする場合，大量の fill-in が発生しメモリ使用量が増大する．そのため，大規模かつ疎な係数行列からなる連立一次方程式を求解する際には，係数行列の変形が伴わない反復解法が用いられている．

反復解法を高速化するにあたり，主要な計算時間を占める疎行列密ベクトル積（Sparse Matrix Vector products：SpMV）を高速化することが求められている．SpMV は行列の各要素に対し 1 回の積和演算のみのメモリ律速な計算である．そのため，CPU と比較して高速なメモリ帯域をもつ GPU（Graphics Processing Unit）を活用することにより高速化が図られてきた．

大規模な疎行列を，GPU の少ないデバイスメモリへ格納するにあたり，メモリ容量効率のよい CSR（Compressed Sparse Row）形式が多く用られている．しかし，CSR 形式での SpMV はストライドアクセスと reduction が必要である．そのため，メモリアクセスパターンを改善し高速に SpMV の計算が可能な SlicedELL 形式や SELL-C-σ 形式 [18] といった疎行列格納形式が提案されている．これらの格納形式では，ストライドアクセスによって発生するメモリアクセスペナルティを減らしデバイスメモリ帯域を引き出しているが，さらなる SpMV の高速化のためにはメモリアクセス回数そのものを減らす必要がある．

メモリアクセスを減らすため，非ゼロ要素の値そのものを圧縮する手法と，非ゼロ要素位置を圧縮する手法があげられる．両者を比較すると，非ゼロ要素の値を圧縮する場合，対象とする疎行列の値が実数やバイナリ，複素数など行列依存の側面が強い．非ゼロ要素位置の圧縮では整数かつ行番号と列番号の組み合わせがユニークなため，値そのものより圧縮しやすい．上記の理由により，本研究では非ゼロ要素位置の圧縮を対象とする．

非ゼロ要素位置を圧縮する手法として差分符号化と辞書圧縮があげられる．差分符号化としては CoAdELL が提案されており，行ベクトルの列番号の差分をとり，ビット数を減らすことによって容量を削減している．しかし，その差分の大きさによってはビット数を減らせず，2 の乗数 bit ではない可変長符号は GPU で計算しにくいという欠点が存在している．一方辞書圧縮方式では対象とする疎行列のパターン性によっては非常に高い圧縮率が期待できる．しかし，GPU は数スレッドを Warp と呼ばれるグループにまとめて，Warp 内の各スレッドが同一の命令を実行する．そのため辞書の単語長が異なると Warp 内のスレッドが実行している命令が異なってしまい，Warp ダイバージェンスと呼ばれる実行効率の低下が発生する．

そのため本論文では，非ゼロ要素位置情報へ辞書圧縮を適用し，メモリへのア

クセスを減らすことによって， GPU 上で SpMV を高速に計算可能とする圧縮疎行列格納形式を提案する．提案手法によって，CSR 形式と比較して最大 29.5%のメモリ使用量を削減し，SpMV の計算時間では最大 19.6%の高速化が得られた．

加えて，これらの SpMV 計算に特化した疎行列格納形式は非ゼロ要素の追加・削除が容易ではないため，より編集が容易な COO 形式で疎行列を生成・保存することが一般的である．そのため，CSR 形式や SELL-C-σ，CoD-SELL を利用するためには格納形式の変換が必要である．特に，CoD-SELL は辞書圧縮のための計算が必要であり，変換時間が大きくなると予測される．そのため，COO 形式から CSR 形式への変換と似た計算時間で変換可能な，CoD-SELL の高速な形式変換を提案する．