



ホワイトペーパー

# NVIDIA GeForce GTX 750 Ti

究極のワットあたりパフォーマンスの実現に向け、第 1 世代 Maxwell  
GPU テクノロジーを採用

V1.0

## 目次

|                                      |    |
|--------------------------------------|----|
| 目次 .....                             | 1  |
| はじめに .....                           | 3  |
| Maxwell の真骨頂: ワットあたりパフォーマンスの向上 ..... | 4  |
| GM107 Maxwell アーキテクチャの詳細 .....       | 5  |
| 次世代の Maxwell SM .....                | 6  |
| メモリ・システム .....                       | 9  |
| 最新のビデオ機能 .....                       | 9  |
| まとめ .....                            | 10 |

## はじめに

最新のハリウッド大作で見られる驚愕の特殊効果や、生きているかのようなキャラクターの登場、写真のようにリアルな 3D ゲーム環境、さらには、高解像度の画像・動画を駆使した、表現力豊かな Web サイトなど、驚愕のグラフィックスに対する一般消費者の要求は、年を追うごとに高まり続けています。このような課題に応えるべく、NVIDIA のグラフィックス・プロセッサも進化を続けており、世代ごとに新たな機能を取り入れることで、より強力な製品へと進化しています。そして、2012 年前半には、画期的なパフォーマンスと電力効率を実現するものとして、NVIDIA の Kepler GPU アーキテクチャが発表されました。Kepler GPU は、世界最速のゲーム用 PC やワークステーション、スーパーコンピュータ、クラウドゲーム・サーバの原動力となりました。また、同アーキテクチャは、Tegra K1 システム・オン・チップ・ファミリにも実装され、スマートフォンやタブレットのほか、車載用のインフォテインメント・システムでも、業界最高のビジュアル・コンピューティング機能を実現しました。

しかし、グラフィックスのリアルさをさらに一段階上に引き上げるため、NVIDIA のエンジニア・チームは、Kepler よりもさらに効率的な次期アーキテクチャを開発する必要性を実感していました。

これを受けて、誕生したのが NVIDIA の第 1 世代「Maxwell」アーキテクチャです。Maxwell には、さまざまな点でアーキテクチャ上の強化が図られており、消費ワットあたりのパフォーマンスと電力効率をこれまで以上に向上できるように設計されています。Maxwell ベースの GPU の第一弾、コードネーム「GM107」は、ノート PC やスモールフォームファクタ(SFF)PC など、電力に制約のある環境での使用向けに設計されています。こうした SFF システムは、最も直近の事例では、Valve Software が発表したばかりの「Steam Machines」イニシアチブなど、ゲームやホーム・エンターテインメントの分野で採用されることが多くなっています。また、GM107 GPU を搭載したグラフィックスカードの第一弾は、GeForce GTX 750 Ti です。GM107 は、アーキテクチャの効率性が極めて優れており、1080p の解像度で GeForce GTX 750 Ti を使用した場合、NVIDIA の 4 年前の主力 GPU、GeForce GTX 480 にほとんど匹敵するパフォーマンスを達成しつつ、その消費電力は同 GPU の 4 分の 1 に相当する、わずか 60W TDP(熱設計電力)ですみます。

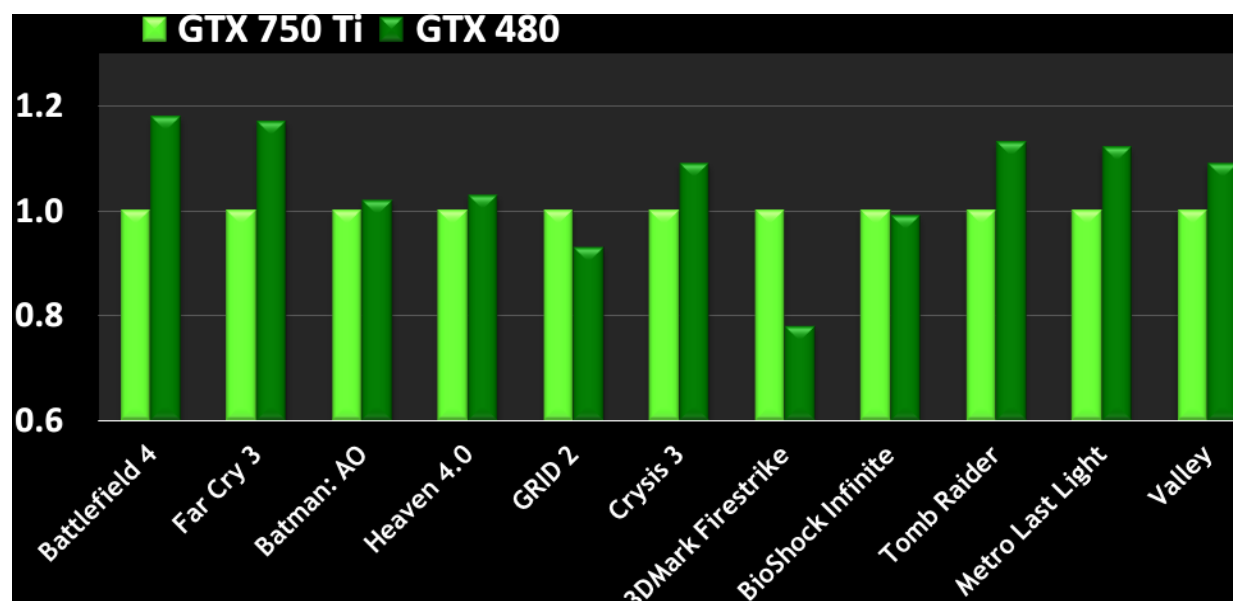


図 1: GeForce GTX 750 Ti、今日の人気タイトルの大半で、GTX 480 に匹敵する動作を実現

GM107 は、第 1 世代 Maxwell アーキテクチャを採用した初の GPU 製品です。これらの第 1 世代 Maxwell 製品は、低消費電力動作に特化しています。また、NVIDIA では、パフォーマンス部門とエンスージアスト・グラフィックス部門を対象とした、より高パフォーマンスの第 2 世代 Maxwell GPU についても、後日の導入を予定しています。

## Maxwell の真骨頂: ワットあたりパフォーマンスの向上

PC やワークステーション、スーパーコンピュータで使用された GPU から、ポケットサイズのモバイル・チップまで、Kepler を移行させる過程の中で、NVIDIA は、GPU の全体的な消費電力の削減だけでなく、同一の電力レベルで、このアーキテクチャからより高いパフォーマンスを引き出す方法についても、多くのことを学びました。そして、こうした取り組みで学んだ内容はすべて、Maxwell に投入されました。

Maxwell では、ストリーミング・マルチプロセッサ (SM) について全く新しい設計が導入され、ワットあたりパフォーマンスと面積あたりパフォーマンスが大幅に高められました。Kepler SMX も、同世代の中では極めて効率的な設計でしたが、その開発過程の中で、NVIDIA の設計チームは、アーキテクチャの効率性を、さらに飛躍的に高める機会を見出しました。Maxwell SM は、こうしたビジョンを具現化したものといえます。そこでは、コントロールロジックの分割、ワークロード・バランシング、クロック・ゲーティングの精度、コンパイラ・ベースのスケジューリング、クロックサイクルあたりの発行命令数の向上など、さまざまな機能強化が導入されており、Maxwell SM (別名「SMM」) は、Kepler SMX の効率性を大きく上回ることが可能になりました。また、新登場の Maxwell SM アーキテクチャによって、NVIDIA では、ダイ面積をわずか 25% 増やすだけで、搭載されている SM の数を、GK107 の 2 基から GM107 では 5 基に増加できました。Maxwell SM で導入された変更箇所については、本書の「次世代 Maxwell SM」の章で詳述しています。

Maxwell では、L2 キャッシュ設計についても大幅な大容量化が図られており、GK107 の 256KB に対し、GM107 では 2048KB となっています。より大容量のキャッシュがオンチップで配置されることで、グラフィックスカードの DRAM に対するリクエストも少なくてすみ、基板の全体的な消費電力を削減しつつ、パフォーマンスを向上できます。

上記の変更内容に加えて、NVIDIA のエンジニア・チームは、Maxwell GPU の各ユニットの実装について、トランジスタの段階まで積極的なチューニングを施すことで、エネルギー効率を極限まで高めています。

こうした取り組みの結果、Maxwell は、Kepler と同じ 28nm の製造プロセスを使用しつつ、Kepler の 2 倍のワットあたりパフォーマンスを実現できます。

## GM107 Maxwell アーキテクチャの詳細

グラフィックス機能の観点から見ると、NVIDIA の第 1 世代 Maxwell GPU は、Kepler GPU と同じ API 機能を備えています。ハイレベルでは、Maxwell は GPC (グラフィックス・プロセッシング・クラスター) 内に複数の SM ユニットを実装しており、各 SM には Polymorph エンジンとテクスチャ・ユニットが、各 GPC にはラスタ・エンジンが搭載されています。これまでと同様、ROP は L2 キャッシュ・スライスおよびメモリ・コントローラと連携しています。内部では、すべてのユニットとクロスバーの構造について設計の見直しが行われており、データフローの最適化や、電力管理の大幅な向上などが実現しています。

GM107 GPU には、1 基の GPC、5 基の Maxwell ストリーミング・プロセッサ (SMM)、2 基の 64 ビット・メモリ・コントローラ (合計 128 ビット) が搭載されています。これは、同チップのフル実装であり、GeForce GTX 750 Ti での出荷時と同じ構成になっています。

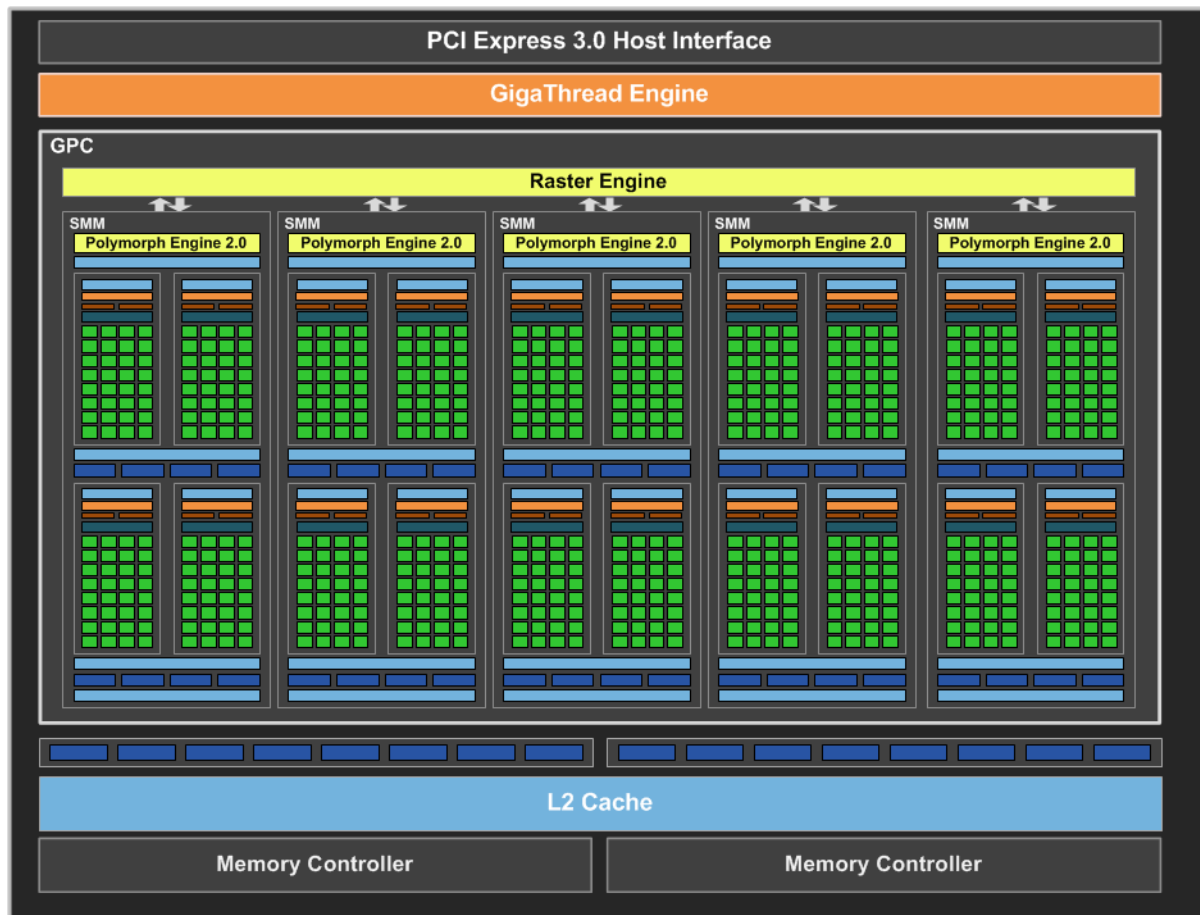


図 2: GM107 のフルチップ・ブロック図

下表は、Maxwell と NVIDIA の旧世代の GK107 Kepler GPU との高レベルでの比較です。

| GPU           | GK107 (Kepler)      | GM107 (Maxwell)     |
|---------------|---------------------|---------------------|
| CUDA コア数      | 384                 | 640                 |
| ベース・クロック      | 1058MHz             | 1020MHz             |
| GPU ブースト・クロック | なし                  | 1085MHz             |
| GFLOP         | 812.5               | 1305.6              |
| テクスチャ・ユニット数   | 32                  | 40                  |
| テクスチャ・フィルレート  | 33.9 ギガテクセル/<br>秒   | 40.8 ギガテクセル/秒       |
| メモリ・クロック      | 5000 MHz            | 5400 MHz            |
| メモリ帯域幅        | 80 GB/秒             | 86.4 GB/秒           |
| ROP           | 16                  | 16                  |
| L2 キャッシュサイズ   | 256 KB              | 2048 KB             |
| TDP           | 64 W                | 60 W                |
| トランジスタ        | 130 億               | 18.7 億              |
| ダイサイズ         | 118 mm <sup>2</sup> | 148 mm <sup>2</sup> |
| 製造プロセス        | 28 nm               | 28 nm               |

[Kepler](#) と [Fermi](#) に関する NVIDIA のホワイトペーパーでは、GigaThread エンジン経由でのホスト PCI Express インターフェイスからのデータフローや、Polymorph およびラスタ・ユニットの基本動作など、アーキテクチャの基本的な側面について詳述しています。これらの動作について、より詳細な背景情報が必要な場合、まずはこれらの文書でご確認いただくことを強くお勧めします。SMM に導入された変更内容については、以下のページで詳述します。

## 次世代の Maxwell SM

Maxwell の効率化に貢献している最も大きな要因は、SMM と呼ばれる、Maxwell の新たな SM アーキテクチャです。この新しい SM アーキテクチャでは電力効率が大幅に高まっており、シェーダバウンドなワークロード上で、CUDA コアあたりのパフォーマンスが 35% 向上しています。こうした結果を実現するには、アーキテクチャに多数の変更を施す必要がありました。SM のスケジューラ・アーキテクチャとアルゴリズムは、より賢く動作して不要なストールを回避しつつ、スケジューリングに必要な命令あたりのエネルギー量をさらに削減できるよう、書き換えが行われました。

さらに、SM の構成も変更されました。各 SM は、4 つの独立した処理ブロックに分割され、それぞれには、専用の命令バッファ、スケジューラ、32 の CUDA コアが搭載されています。Kepler では、2 の累乗ではないコア数の CUDA を使用し、その一部を共有するというアプローチを採用していましたが、これ

は廃止されました。こうした分割によって、設計とスケジューリング用の論理回路が簡素化されており、面積と消費電力を抑え、演算の遅延を低減できます。

また、処理ブロックをペアにして、4つのテクスチャ・フィルタリング・ユニットとテクスチャ・キャッシュを共有しています。今回、コンピュータ L1 キャッシュ機能は、テクスチャ・キャッシュ機能と統合されており、共有メモリは、独立したユニット(初の CUDA 対応 GPU である G80 で使用されたアプローチと類似)となっており、4つのブロックすべてで共有されています。

全体的には、この新たな設計によって、各「SM」は、大幅な小型化を実現しつつ、Kepler SM の約 90% のパフォーマンスを実現しています。また、省面積化によって、NVIDIA は、GPU あたりでより多くの SM を実装できます。GM107 全体の SM 関連の指標を GK107 と比較すると、GM107 は、GK107 の 2 基に対して 5 基の SM を搭載しており、ピーク時のテクスチャ性能は 25% 高く、CUDA コア数は 1.7 倍で、シェーダ性能は約 2.3 倍となっています。

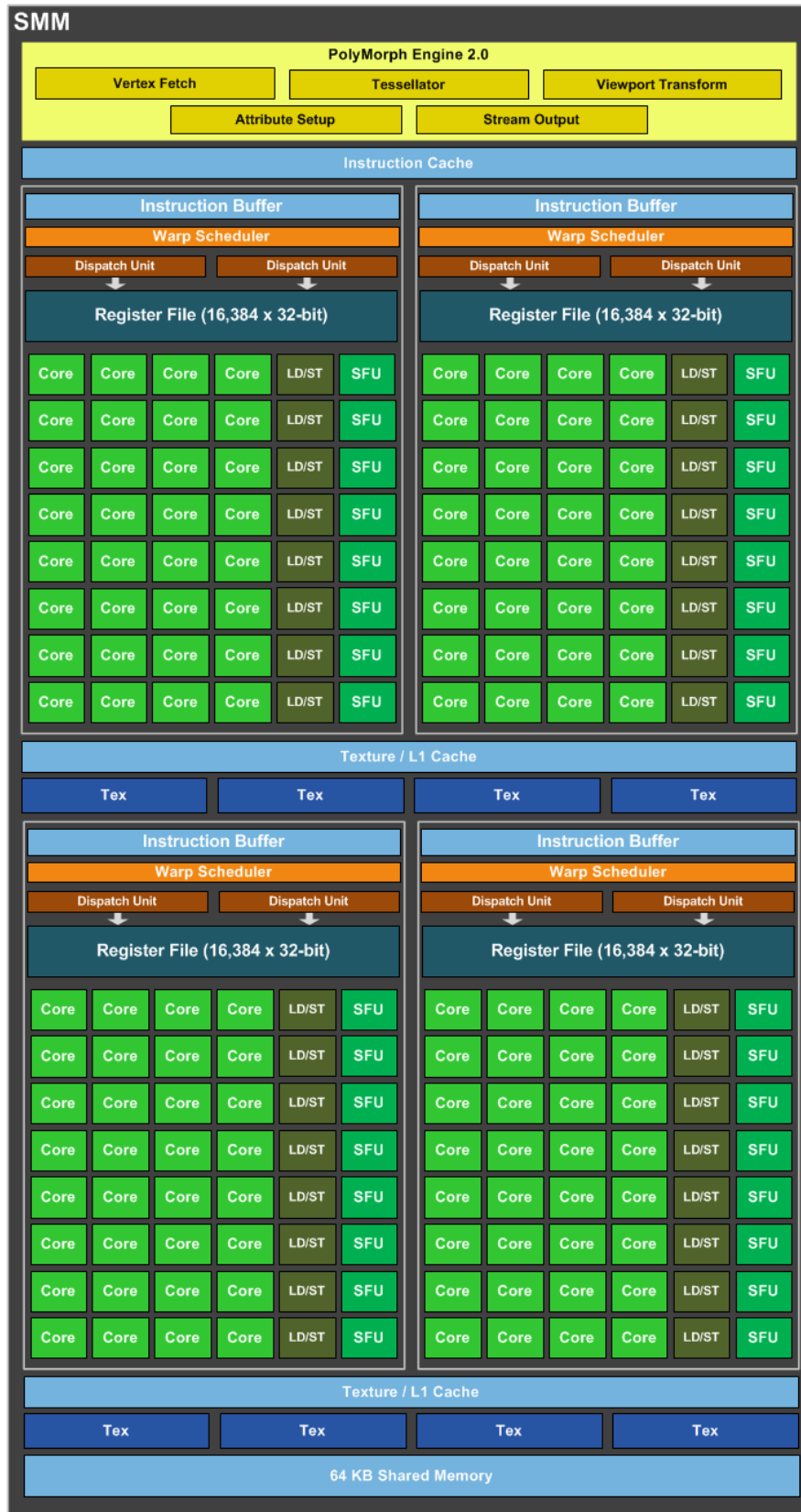


図 3:Maxwell SM のブロック図

## メモリ・システム

GK107と同じメモリ幅を使用しつつ、GK107を大幅に上回るパフォーマンスを実現するため、GM107では、メモリ・システムの機能強化に対する投資も重要でした。また、設計の効率化と平行して、内部のメモリ・システムの帯域幅も上昇しました。さらに、大容量な2MB L2 キャッシュ構成(従来のどのGPU設計よりも大容量)により、必要なメモリ帯域幅を抑えつつ、DRAM帯域幅がボトルネックとなる状態を回避する点について、大きな効力を発揮しています。

## 最新のビデオ機能

これまでの GeForce GPU から進化した、Kepler の主要なイノベーションの1つとして、ハードウェア・ベースによる H.264 ビデオ・エンコーダ、NVENC の存在がありました。(GeForce GPU の CUDA コアを使用するのではなく)エンコード/デコード用の専用のハードウェア回路を集積することで、NVENC は、H.264 エンコードの大幅な高速化と、消費電力の削減を両立できたのです。

そして、NVIDIA は昨秋、Kepler の NVENC エンコーダを活用して、GeForce GTX 600 シリーズと GTX 700 のゲーマー向けに ShadowPlay を導入しました。これによって、ゲーマーは、プレイ中のお気に入りの瞬間を記録し、第三者に公開できるようになりました。ShadowPlay の提供開始以降、300 万件以上の動画が記録されており、ゲーマーは、YouTube への投稿や、Twitch によるプレイ動画のライブ・ストリーミングを楽しんでいます。

そして今回、Maxwell では、ビデオ性能を向上させるために NVENC ブロックが強化されており、エンコードの高速化(Kepler の4倍速に対し、H.264で6~8倍速)とデコードの8~10倍の高速化が実現しています。また、ローカル・デコーダ・キャッシュを新たに導入し、動画のデコード時のストリームあたりのメモリ効率が向上した結果、動画のデコード時の消費電力も抑えられています。

Maxwell には、動画の再生などの軽負荷時の GPU の電力消費を抑えることに特化した、新登場の GC5 パワーステートも採用されています。GC5 は、低消費電力のスリープ状態であり、これまでの GPU 製品と比べて、こうした使用事例での消費電力を大幅に削減できます。

## まとめ

小型化をさらに推し進めた半導体の製造プロセスノードの開発における課題の高まりを踏まえ、NVIDIA のエンジニア・チームは、PC グラフィックスを進化させ続けるには、Maxwell アーキテクチャはより効率化が進んだものにならないといけないことをいち早く実感していました。これは単に、Kepler を大型化すればすむ、という類のものではありませんでした。むしろ、Maxwell では、画期的なワットあたりパフォーマンスを実現することを目指しました。

パフォーマンスを高めつつ、電力の無駄を最小限に抑えるため、SM は 4 つにグループ分けされ、各グループは、スケジューリングと命令ディスパッチ用の専用のリソースを確保しています。また、L2 キャッシュの大幅な大容量化によって、テクスチャのリクエストやアトミック操作などにおいて増加分のストレージ・バッファを GPU 全体で共有することが可能で、メモリへのアクセスを抑制できます。

Maxwell の最新 SMM に施された変更内容により、GPU のハードウェア・ユニットがより高頻度に利用されることで、パフォーマンスと電力効率が向上しています。GeForce GTX 750 Ti は、GK107 との比較で 1.7 倍以上のパフォーマンスがありながら、TDP はわずか 60W です。

今から 20 年前の PC は、もっぱらホームオフィスにある 2 フィート (約 60cm) のタワー型筐体に閉じ込められていました。しかし現在、PC は家庭内の至る所で見ることができ、その寸法は、シューズボックスほどにまで小さくなっています。しかし、こうした市場のユーザの大半は、CPU と内蔵グラフィックスの構成で妥協することが一般的で、ゲームの設定はフレームレートが「低」、グラフィックスが「超低」となるため、小型 PC でのゲームプレイは、少々ありきたりな体験となっていました。しかし、圧倒的な電力効率を誇る Maxwellのおかげで、ホームシアターPC やその他のスモールフォームファクター PC を使用するゲーマーも、解像度 1080p での素晴らしいゲーム体験について、妥協の必要はなくなりました。

すなわち、GeForce GTX 750 Ti は多くの種類の PC に簡単に装着することができます。大半の使用環境では、電源のアップグレードに関する心配は不要です。その結果、市販のベーシックな PC を、強力なゲーム用 PC へと一変させることが可能です。そして、GeForce GTX 750 Ti は、消費電力が極めて低いことから、騒音や発熱も非常に低く抑えられており、ホームシアターPC での使用に最適です。本製品は、電源コネクタの不要な世界最速のグラフィックスカードです。

NVIDIA は、ワットあたりパフォーマンスに対して、極めて積極的な取り組みを行っており、Maxwell は、世界で最も効率的な GPU となっています。その結果、従来考えられなかったようなフォームファクターの PC でも、お気に入りのゲームをプレイすることが可能になりました。

#### **注記**

解説、見解、NVIDIA のデザイン仕様、リファレンスボード、ファイル、図面、診断、リスト、その他のドキュメントなど、本ホワイトペーパーで提供する情報は(以下、集会的あるいは個別に「マテリアル」と呼称します)、すべて、「現状のまま」を条件に提供するものです。NVIDIA は、本マテリアルについて、明示的、暗示的、法定的を含む保証を一切行わず、また、権利非侵害、商品性、および特定目的適合性に関するあらゆる黙示保証を明示的に否認します。

記載された情報の正確性、信頼性には万全を期しておりますが、これらの情報の使用の結果に対して、もしくはこれらの情報の使用に起因して第三者の特許またはその他の権利の侵害が発生しても、NVIDIA Corporationは一切の責任を負わないものとします。暗示的か明示的かを問わず、NVIDIA Corporationの特許または特許権に基づいて付与されるライセンスは一切ありません。本文書に記載された仕様は予告なしに変更されることがあります。本文書は、過去に提供されたすべての情報に優先し、これに代わるものとします。NVIDIA Corporationの製品は、NVIDIA Corporationの書面による明示的な許可なくしては、生命維持装置の重要な部品として使用してはなりません。

#### **商標について**

NVIDIA、NVIDIAのロゴ、CUDA、FERMI、KEPLER、MAXWELLおよびGeForceは、米国およびその他の国におけるNVIDIA Corporationの商標または登録商標です。その他の会社名および製品名は各社の商標である可能性があります。

#### **Copyright**

© 2014 NVIDIA Corporation. All rights reserved.