



TCP転送を40GbEの限界で実現

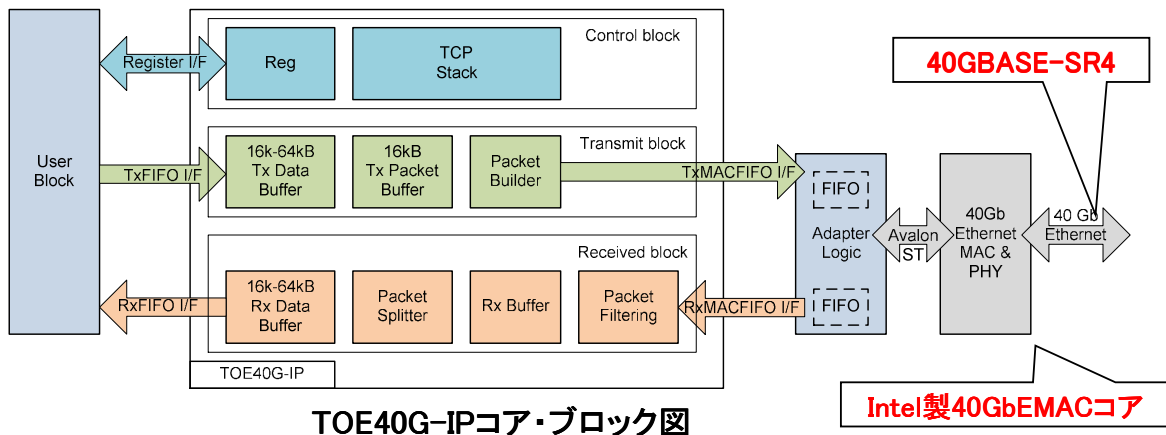
2018/12/14

Design Gateway

Page 1

TOE40G-IPコアの概要

- ・ 40GBASE-SR4対応のTCP/IPオフロード・エンジン・コア
- ・ ユーザ回路とIntel製EMACコアの間に挿入
- ・ TCPの送信・受信の両方向の処理を完全HW化
- ・ Full Duplex(送受信同時)通信をサポート



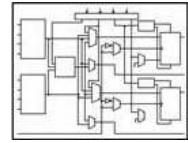
2018/12/14

Design Gateway

Page 2

TOE40G-IPコアの特長1

- ・ TCP/IP処理を完全ロジック・ハードウェア化
 - CPU負荷がゼロ、理論上最小の処理レイテンシ
- ・ 送信のみ/受信のみ/同時送受信を高速転送
 - 4Gyte/sec以上の実転送速度(ジャンボ・フレーム時)
 - 同時送受信(全二重通信)でも同等の速度を維持
- ・ 転送データの信頼性を確保
 - 送信時:ACK抜け/重複/タイムアウト等で自動リトライ
 - 受信時:シーケンス番号を評価し返信ACKを自動制御



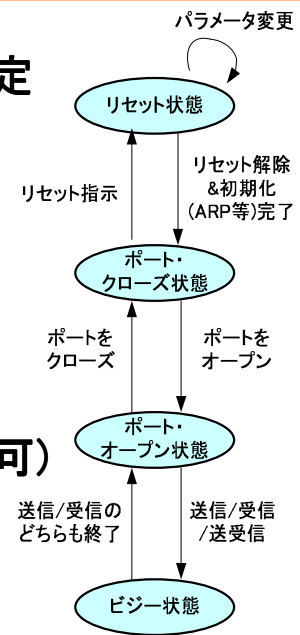
TOE40G-IPコアの特長2

- ・ データバッファ容量を選択可能
 - FPGAメモリ・リソースとパフォーマンスからユーザが選択
- ・ Intel製40GEMACコア(低レイテンシ40Gb MAC)対応
 - 40GBASE-SR4で実機検証済み,他の環境は要問合せ
- ・ 実機動作リファレンス・デザインを用意
 - Intel評価ボードで動作するプロジェクト
 - 購入前にパフォーマンス・信頼性の実機評価を検証可能
 - 製品のリファレンスはコア以外の全回路をソースで添付



TOE40G-IPコアの動作概要

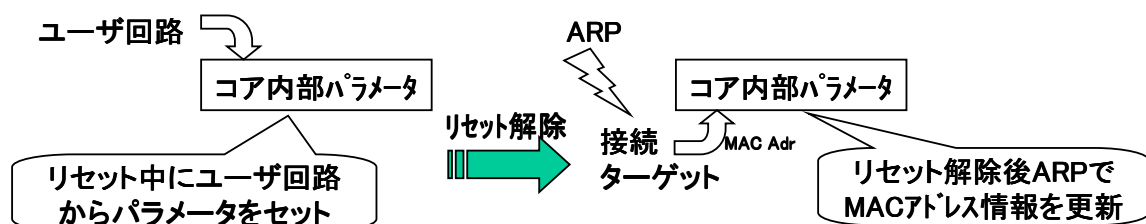
- ・ リセット状態でパラメータ(IP&MACアドレス等)を設定
- ・ リセット解除で初期化(ARP等)を実行
- ・ 初期化完了後ポートをオープン(アクティブ/パッシブ)
- ・ 送信・受信は独立して動作(同時送受信可)
- ・ パラメータ変更はリセット状態で実施
(転送長/パケット長はビジー以外の状態で変更可)



コアの状態遷移図

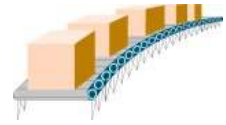
初期化動作

- ・ パラメータの初期値設定
 - コアのリセット維持中にユーザ回路より設定
 - IPおよびMACアドレス・ポート番号を指定
 - 設定を完了するとリセットを解除
- ・ リセット解除後ARP実行
 - クライアント・モード: 接続ターゲットに対してARPを発行
 - サーバー・モード: 接続ターゲットからのARPを待機



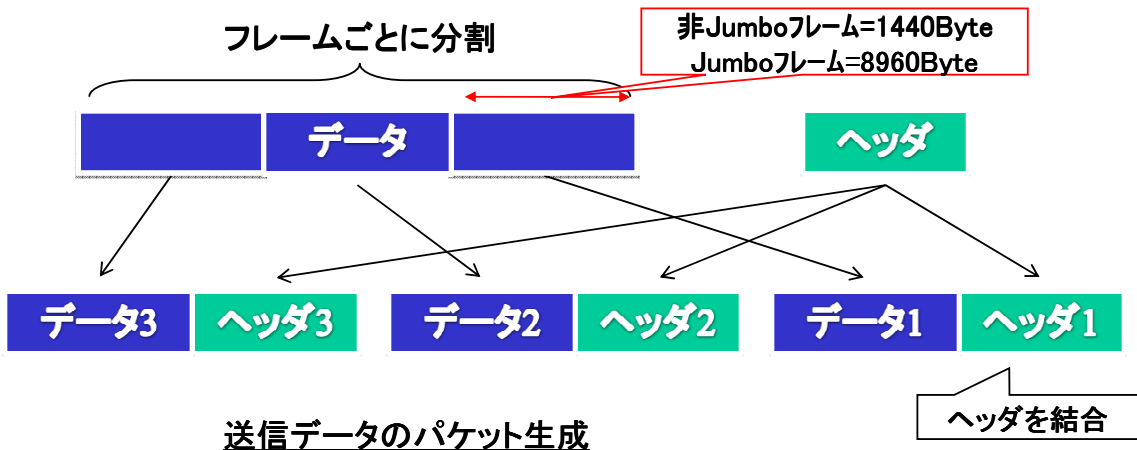
高速送信

- ・ 送信パケットの生成
 - ユーザ回路は送信データをFIFO I/Fで書込み
 - 送信データをフレームサイズで分割
 - ヘッダと送信データを結合し40GEMACへ出力
- ・ 自動再送機能
 - ターゲットからのACKを常時チェック
 - ACK抜け/重複/タイムアウト等の異常ACKを検出
 - 異常ACKの種類に応じて適切な再送データを用意



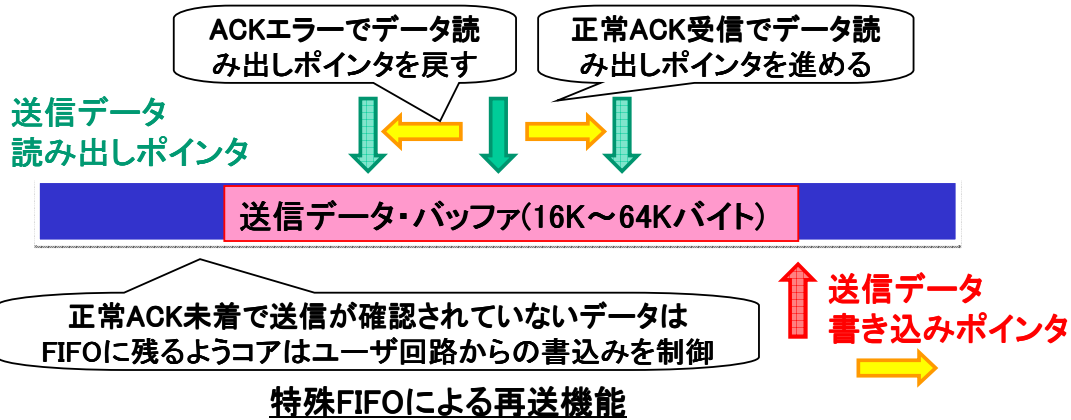
送信パケットの生成

- ・ 送信データにヘッダを自動的に付加しEMACへ転送
 - データはコア内にてフレーム・サイズで分割
 - チェックサムやシーケンス番号などもコアが自動生成






自動再送機能

- 専用設計された特殊なFIFOで再送を実装
 - 正常ACKで読み出しポインタを進める
 - 異常ACKの場合コアは適切な位置にポインタを戻す
 - ポインタ制御・再送の実行をコアが自動制御



高速受信

- 受信パケットのヘッダ・チェック 
 - 受信対象でない場合やチェックサム・エラーの場合は破棄
- データ並び替え 
 - シーケンス番号の入れ替え発生時に並び替え
 - 受信済みデータの再送要求を防ぎ転送効率を維持
 - 並び替え処理できない場合は重複ACKで再送要求
- 重複データの結合 
 - 受信データが前回受信パケットと重複する部分を検出
 - 重複箇所のみ破棄し連続データとして復元

受信パケットのヘッダ・チェック

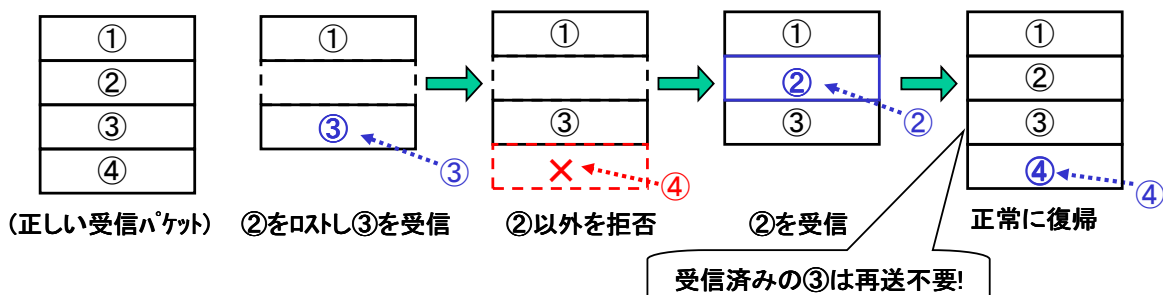
- ・ 受信ヘッダのチェック・サムが正しいことを確認
 - 更に以下の条件を満たすこともコアが確認

Byteオフセット	プロトコル	データ内容	確認条件
0-5	ICMP	あて先MACアドレス	SML/SMHLレジスタで設定したコアのMACアドレスと一致
6-11	ICMP	送信元MACアドレス	ARPで検出した通信ターゲットのMACアドレスと一致
12-13	ICMP	タイプ	0x0800 (IPパケット)であること
14	IP	バージョン/ヘッダ長	0x45 (IPv4, IPヘッダ長=20)であること
20	IP	フラグ/フラグメントOFS	下位6ビットがゼロ(フラグメントの有無チェック)であること
23	IP	プロトコル番号	0x06(TCPパケット)であること
26-29	IP	送信元IPアドレス	DIレジスタで設定した通信ターゲットのIPアドレスと一致
30-33	IP	あて先IPアドレス	SIレジスタで設定したコアのIPアドレスと一致
34-35	TCP	送信元ポート番号	DPNレジスタで設定/パッシブオープンで取得した通信ターゲットのポート番号と一致
36-37	TCP	あて先ポート番号	SPNレジスタで設定したコアのポート番号と一致
38-41	TCP	シーケンス番号	前回処理時のシーケンス番号との差分がコアで処理可能範囲内

受信パケット・ヘッダの確認条件

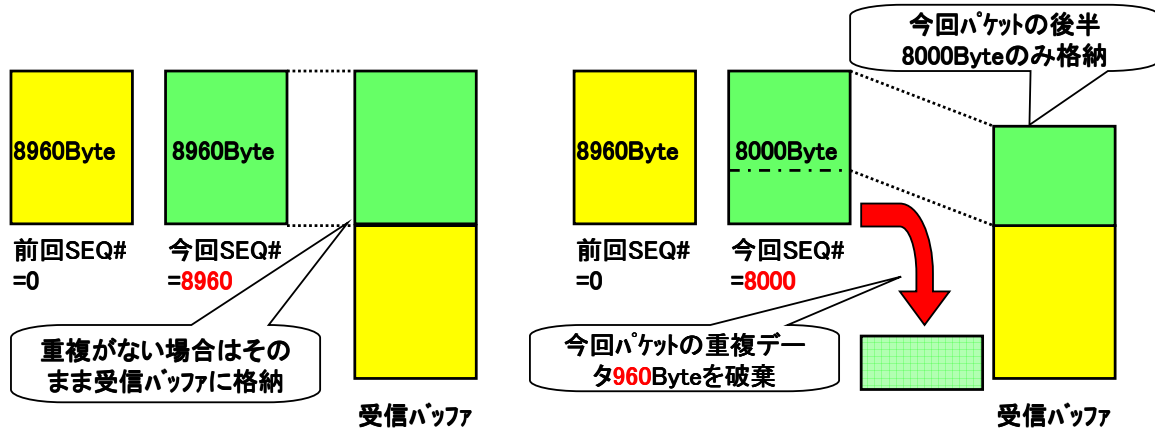
データ並び替え

- ・ シーケンス番号がスキップした(ロスト)場合に機能
 - ロスト状態を解消するパケット以外は受信しない
 - スキップ部のデータを回復するロスト解消パケットのみ受信
- ・ データ並び替え
 - ロスト解消パケットからデータの連続性を完全に復元
 - 受信済みデータは再送要求せずパフォーマンスを維持



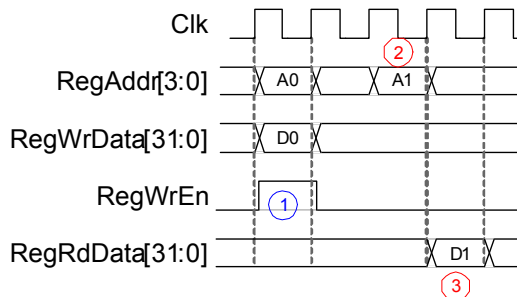
重複データの結合

- データの重複を検出し自動的に補正
 - シーケンス番号から受信データの重複を検出
 - 重複箇所のみ破棄し連続データとして復元



ユーザ・インターフェース(制御)

- レジスタI/F、送信FIFO I/F、受信FIFO I/Fの3種類
 - レジスタI/Fは初期パラメータの設定、方向切り替え指示
 - 送信データ・受信データ用I/Fは標準的なFIFO I/F

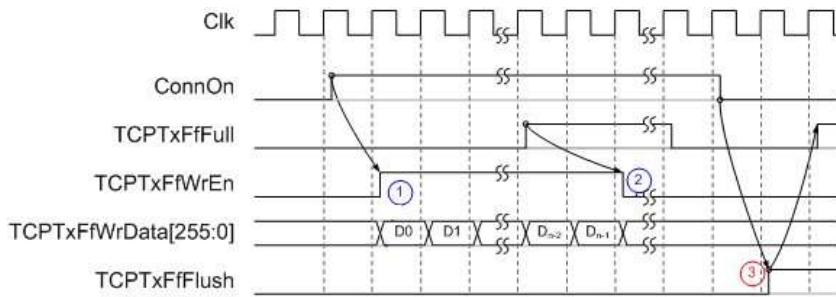


[レジスタの書込み]
①アドレスとデータを
設定しWrEnで書込み

[レジスタの読出し]
②アドレスを与えたる
③次クロックで有効
データが出力

レジスタI/Fのタイムチャート

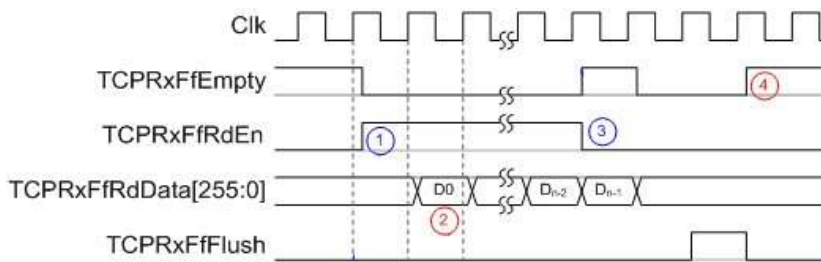
ユーザ・インターフェース(データ)



送信FIFO I/Fのタイムチャート

[送信データの書込み]

- ①データをWrEnで書込み
- ②Fullになってから4クロック以内にライト中断
- ③FlushでFIFOクリア



受信FIFO I/Fのタイムチャート

[受信データの読み出し]

- ①非EmptyでRdEnにて読出し
- ②次のクロックでデータ出力
- ③Emptyではリード禁止
- ④FlushでFIFOクリア

データ・バッファ容量の設定

- ・ 送受信のデータ・バッファをパラメタライズで設定可能
 - ① 送信データ・バッファ: 16K/32K/64KByteから選択
 - ② 受信データ・バッファ: 16K/32K/64KByteから選択
- ・ リソースとパフォーマンスの最適点を調整できる

ジェネリック名	設定範囲	説明
TxBufBitWidth	9-11	送信データ・バッファ・サイズをアドレス・ビット幅で設定します。例えば 9 の場合 16K バイト、11 の場合 64K バイトとなります。
RxBufBitWidth	9-11	受信データ・バッファ・サイズをアドレス・ビット幅で設定します。例えば 9 の場合 16K バイト、11 の場合 64K バイトとなります。

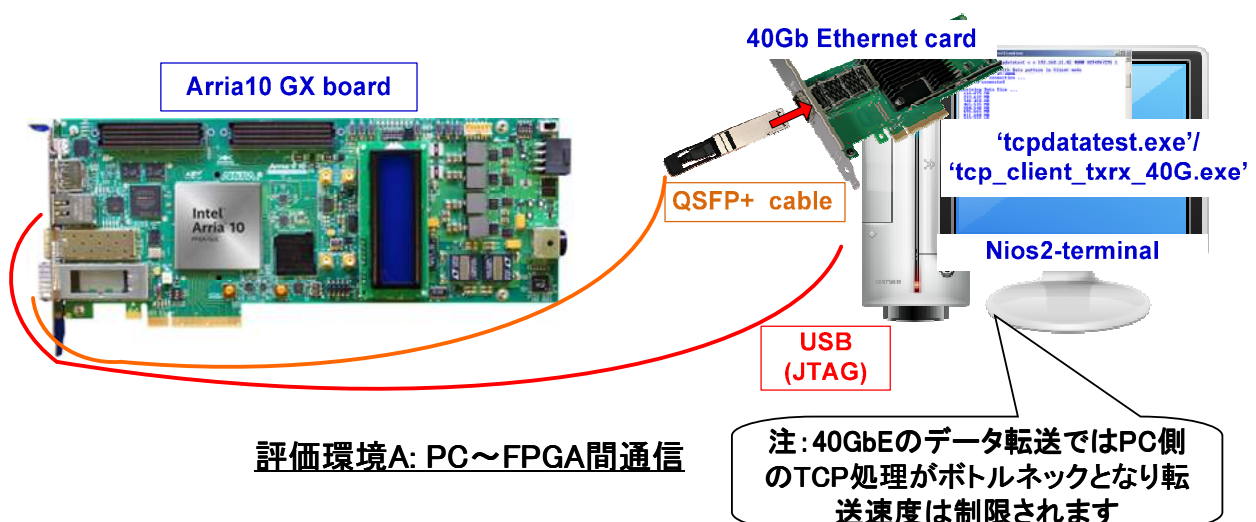
各データ・バッファはパラメタライズで設定できる

評価用SOFファイル

- ・ 2種類の評価用bitファイル
 - PC～FPGA間または2枚のFPGAボード間で通信
 - ただしPC評価の場合PC側がボトルネックになる
(汎用PCで40Gbit/secのTCPデータ実時間処理は困難)
- ・ 転送パフォーマンス測定/データ・ベリファイ
 - 半二重通信および全二重通信をサポート
 - ベリファイによりデータ信頼性の確認が可能
- ・ SOFファイルでコア購入前に実機評価可能

評価環境A: PC～FPGA間通信

- ・ PC側は評価用ソフトウェアを使用し評価
 - 注意点:PC側がボトルネックで最高パフォーマンスは出せない



リファレンスと実機評価による開発

- ・ リファレンス+評価ボードによる確実な開発
 - まず最初に製品添付のリファレンスで実機動作を確認
 - そこからユーザ製品に向け少しずつ編集
 - 編集ごとに実機動作をStep by Stepで確認
 - 問題があれば1ステップ前に戻るだけで動く状態にすぐ復帰できる



大きな後戻りがなく確実に短期間での製品開発が可能!

消費リソース

- ・ Arria10GXシリーズのコア単体消費リソース
 - 送信/受信ともデータ・バッファを最大の64Kバイトに設定した場合のリソース



ファミリ	ターゲットデバイス	Fmax (MHz)	ALMs ¹	Registers ¹	Block Memory bit ²	Design Tools
Arria 10 GX	10AX115S2F45I1SG	322	3,656	5,696	1,179,648	QuartusII18.0

TOE40G-IPコア単体コンパイル結果

送受信ともバッファを最大に設定した場合です。
バッファ容量を削減すればBlockメモリ消費リソースを節約できます。

転送パフォーマンス(半二重通信)

```

+++ TOE40G-IP Send Mode +++
Enter transfer size (aligned to 256-bit): 32 - 0x1FFFFFFE0 => 0x1FFFFFFE0
Enter packet size (aligned to 256-bit) : 32 - 8960 => 8960
Input mode : [0] Client [1] Server => 1
Wait Open connection ...
Connection opened
Start data sending
Send 4957 MByte Recv 0 Byte
Send 9913 MByte Recv 0 Byte
|
|
Send 128880 MByte Recv 0 Byte
Send 133837 MByte Recv 0 Byte
Send data complete

Close connection
Connection closed

Total tx transfer size = 4294967295 (256-bit)
Total = 137.438[GB]
Time = 27726[ms] , Transfer speed = 4907[MB/s]
    
```

128Gバイト・データを
ジャンボフレーム(8960バイト)
で他方のFPGAへ送信

**転送パフォーマンス
4907MByte/sec!**

2枚のFPGAボード間での片方向(送信→受信)評価結果

転送パフォーマンス(全二重通信)

```

+++ TOE40G-IP Full-duplex Mode +++
Enter total size (aligned to 256-bit) : 32 - 0x1FFFFFFE0 => 0x1FFFFFFE0
Enter packet size (aligned to 256-bit): 32 - 8960 => 8960
Enable data verification : [0] Disable [1] Enable => 1
Input mode : [0] Client [1] Server => 0
Open connection
Connection opened
Start data transferring
Send 4939 MByte Recv 4931 MByte
Send 9878 MByte Recv 9837 MByte
|
|
Send 133355 MByte Recv 132505 MByte
Send 137438 MByte Recv 137419 MByte
Send data complete

Close connection
Connection closed

Total tx transfer size = 4294967295 (256-bit)
Total = 137.438[GB] , Time = 28003[ms] , Transfer speed = 4858[MB/s]

Total rx transfer size = 4294967295 (256-bit)
Total = 137.438[GB] , Time = 28003[ms] , Transfer speed = 4858[MB/s]
Input any key to stop the test
    
```

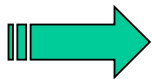
128Gバイト・データを
ジャンボフレーム(8960
バイト)で同時送受信

**転送パフォーマンス
4858MByte/sec!**

2枚のFPGAボード間での双方向(同時送受信)評価結果

結論

- ・ TOE40G-IPコアで接続ターゲットと高速データ転送
 - 保証されたTCPデータで4GByte/secの転送速度
- ・ 再送を含めた複雑なTCP処理を完全にハード化
 - データ並び替えや重複データ削除もコアが自動的に処理
- ・ レジスタやFIFOにより接続容易なユーザI/F
 - CPUなしのシステム実装を可能とするシンプルなI/F
- ・ 実機動作を確認できるリファレンス・デザイン
 - デザインと動作確認の同時並行が可能



**40GbEメリットを最大限に活用する
製品開発を強力に支援します!**



2018/12/14

Design Gateway

Page 25

問い合わせ

- ・ ホームページに詳細な技術資料を用意
 - https://dgway.com/TOE40G-IP_A.html
- ・ 問い合わせ
 - 株式会社Design Gateway
 - E-mail : info@dgway.com
 - FAX : 050-3588-7915



2018/12/14

Design Gateway

Page 26



改版履歴

Rev.	日時	履歴
1.0J	2018/10/4	日本語プレゼン初版リリース