

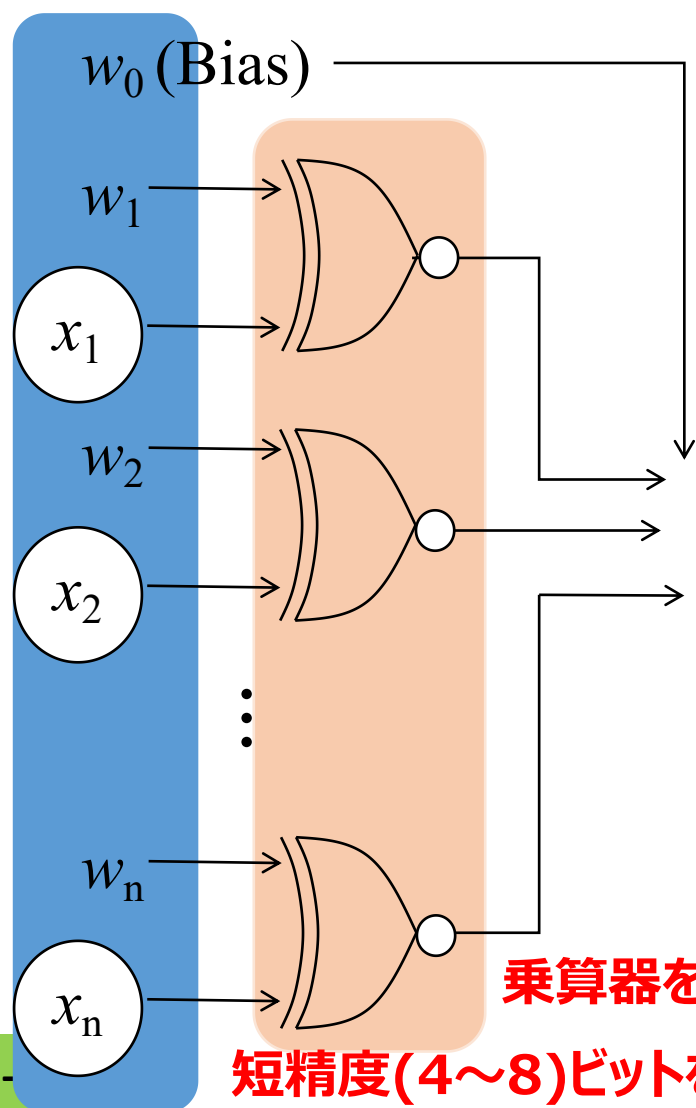
本スライドは、当日のセミナー資料の一部を抜粋したものです。

組込み(エッジ)でディープラーニング

- クラウドでの問題
 - ネットワーク遅延
 - プライバシー
 - セキュリティ
- 学習はオンライン,
推論だけ行うことを想定
- 検討事項
 - 計算能力
 - バッテリー
 - 冷却ファン
 - バッテリー時間



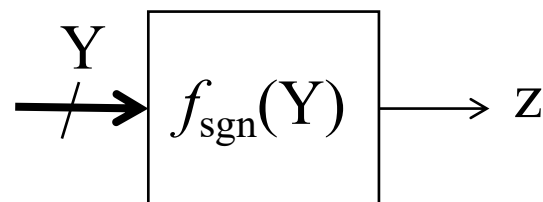
2値化ニューラルネットワーク



x1	x2	Y
-1	-1	1
-1	+1	-1
+1	-1	-1
+1	+1	1



x1	x2	Y
0	0	1
0	1	0
1	0	0
1	1	1



乗算器をXNORに置き換え→回路面積の削減

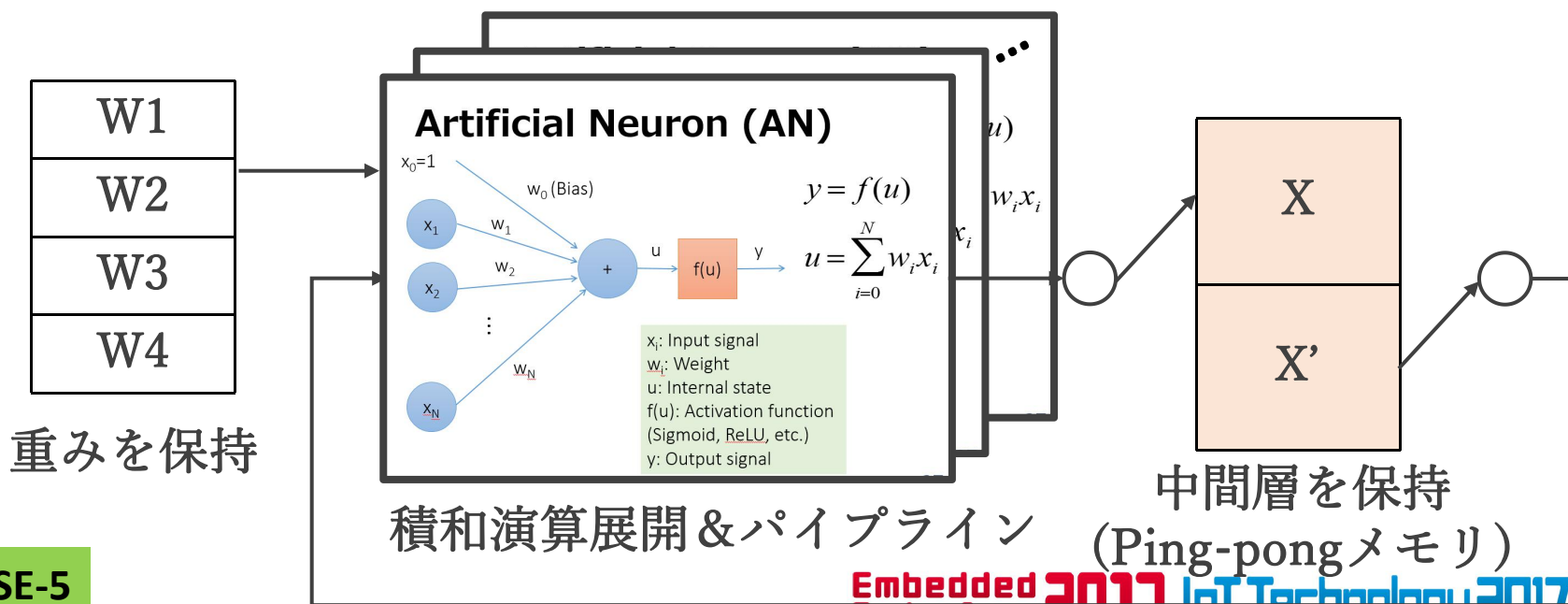
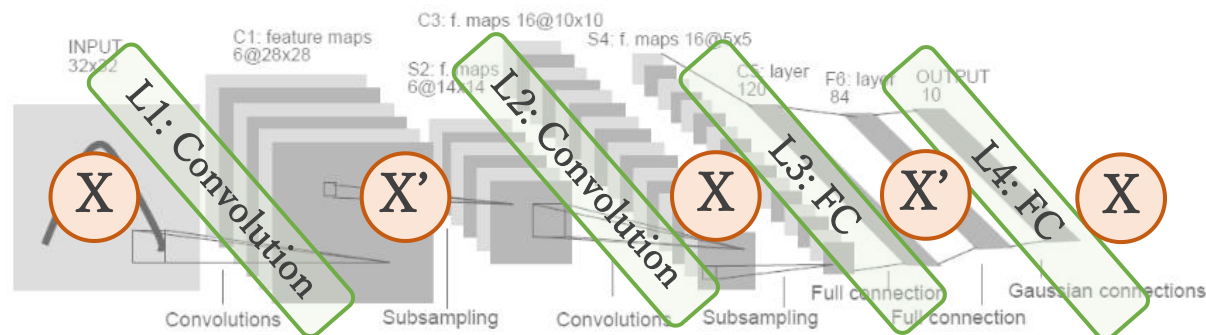
短精度(4~8)ビットを2値に置き換え→メモリ帯域の圧縮

2値化と既存ディープラーニングの比較

項目	既存 ディープラーニング	2値化 ディープラーニング
計算精度	64ビット 浮動小数点	1ビット 論理値
畳み込み演算方式	算術演算	論理演算
ターゲット分野	クラウド (数1000クラス分類)	組み込み機器 (数10クラス分類)
プラットフォーム	CPU, GPU	FPGA, カスタムチップ

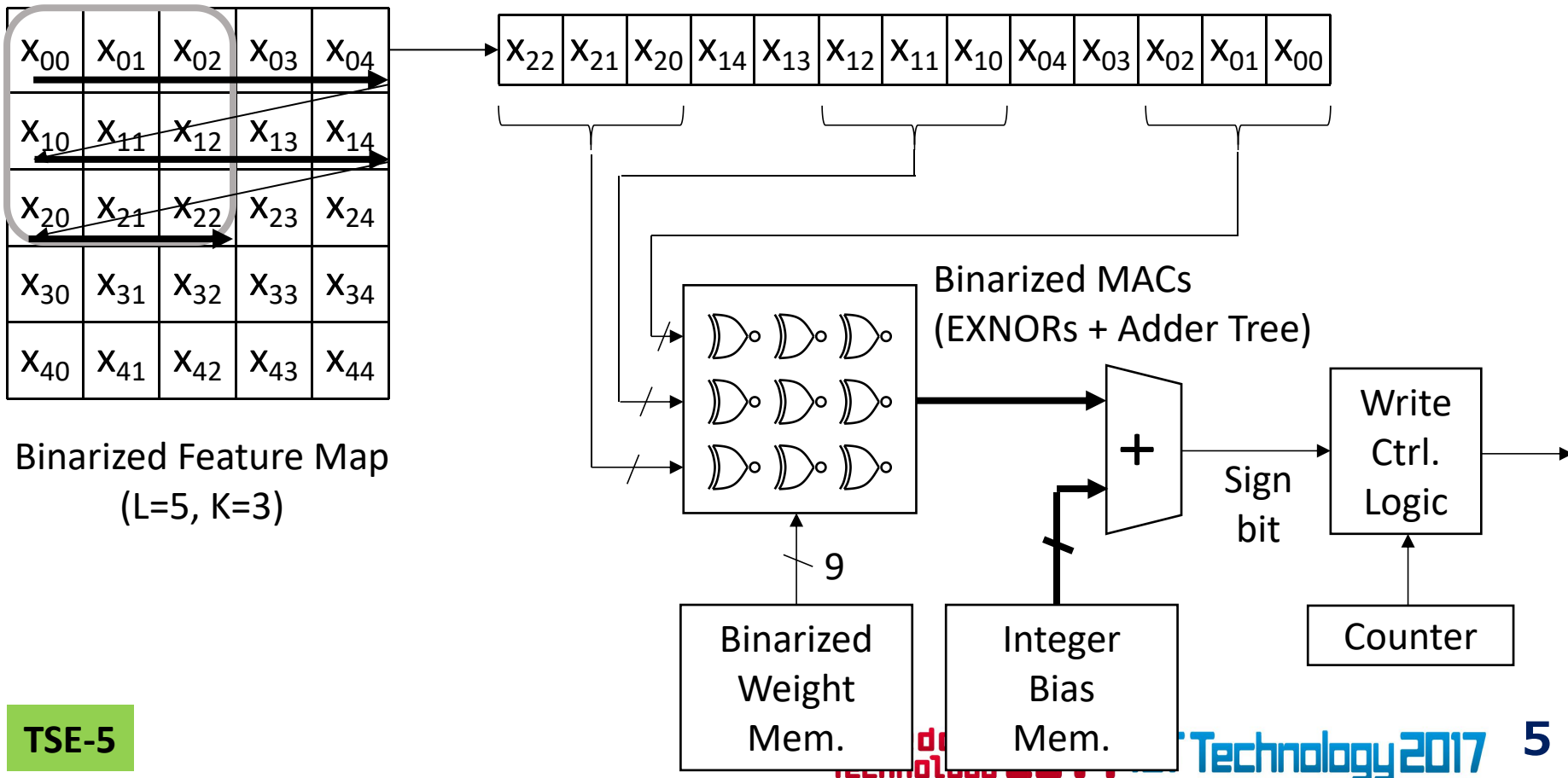
FPGA実装の全体像

各層を時分割で実行, 層毎にパイプライン処理

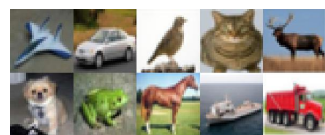


2値化CNN専用回路

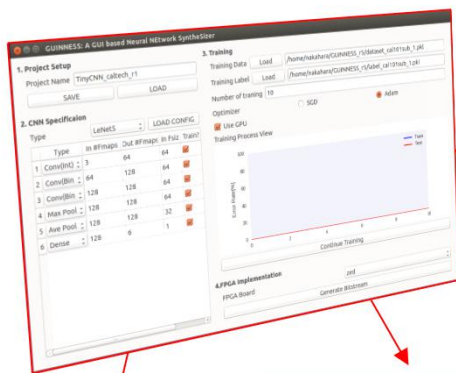
- カスタマイズ演算: 1ビット積和演算
- 専用パイプライン



GUINNESSツールフロー



Generated from
Images



Operated by
the GUI

