

ニューロモルフィック イベントベースセンサーの基礎

目次

序論	1
時間分解能とは	2
時間分解能はいつでも高いほど良いのか?	3
非同期イベント読み出しスキームと 同期イベント読み出しスキームの長所と短所は何か	8
ダイナミックレンジ及び低照度性能はどのようにして決まるか?	16

序論

ニューロモルフィックイベントベースビジョンは、急速に拡大している新しい分野である。本技術によって、レイテンシ(応答時間)、時間分解能、及びダイナミックレンジといったセンサー性能のさまざまな面でメリットが期待される。ただし、標準的試験方法が存在しないので、この分野のさまざまなプレーヤーからの評価を比較し検証するのは、現在のところ困難である。加えて、さまざまな使用事例で各性能面の相対的な重要性についても、議論が続いている。本概説では、ニューロモルフィックイベントベース視覚センサーの性能仕様の解釈に関して、重要な問題を質疑応答形式で考察する。

時間分解能とは？

ニューロモルフィックイベントベース視覚センサーの**時間分解能**は、イベント検出時間の個々の測定分解能と定義される。時間分解能は、イベントタイムスタンプの時間単位として表される。イベントタイムスタンプの時間単位は小さいほど、高い時間分解能に相当する。

時間分解能はいつでも高いほど良いのか？

時間分解能の有用性は、タイムスタンプスキームの設計とピクセルフロントエンドノイズにより制限される。

ピクセルフロントエンドノイズの発生源には、フォトンショットノイズ、回路温度ノイズ、及び固定パターンノイズがあり、これらはピクセル(画素)全体を通した時間コントラスト検出閾値におけるミスマッチに起因する。これらのノイズ発生源はすべて基礎物理によって定義されるもので、大部分は不可避である。このノイズはイベント検出時間における不確実性と不均一性をもたらし、図1に示すように、標準照度条件で通常は数百マイクロ秒(μs)のオーダーである。

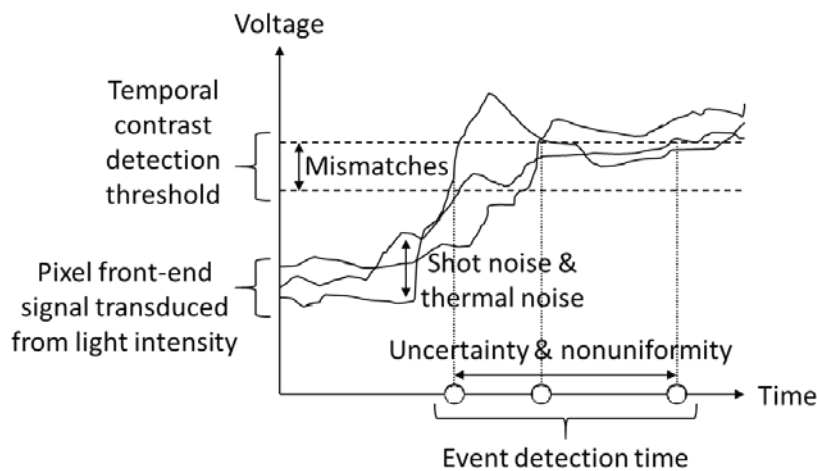


図1) イベント検出時間における不確実性と不均一性

Voltage	電圧
Temporal contrast detection threshold	時間コントラスト検出閾値
Mismatches	ミスマッチ
Pixel front-end signal transduced from light intensity	光強度から変換されたピクセルフロントエンド信号
Shot noise & thermal noise	ショット&温度ノイズ
Uncertainty & nonuniformity	不確実性&不均一性
Time	時間
Event detection time	イベント検出時間

一例として、あるシーンにおいて剛体端がピクセルアレイ(画素列)を横切って移動している投影を考えてみよう。図2に示すように、これは剛体端の投影を受け取るピクセル(これを投影ピクセルと呼ぼう)の光強度を変化させる。

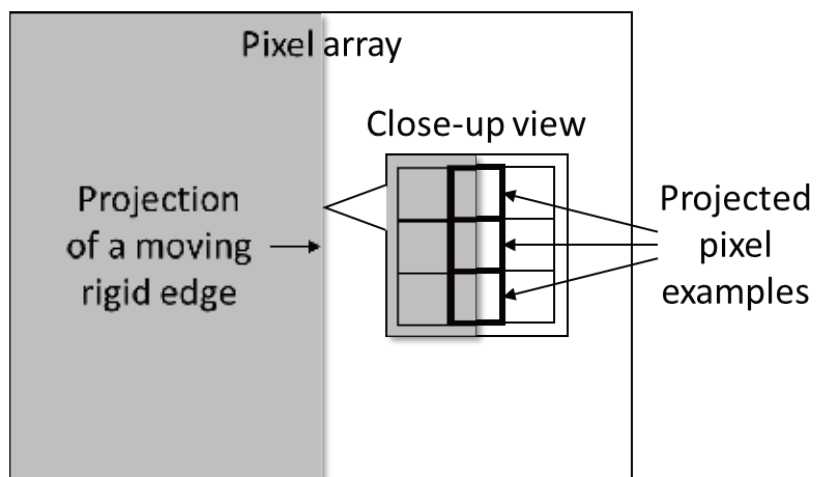


図2) 入力刺激の例

Pixel array	ピクセルアレイ
Projection of a moving rigid edge	移動する剛体端の投影
Close-up view	拡大図
Projected pixel examples	投影されているピクセルの例

グラウンドトゥルースの入力刺激を用いれば、どの瞬間でも、全ての投影ピクセルが同時に光強度の変化を示す。なぜならこれらの変化は、単一の剛体端の動きが引き起こすからだ。しかし、ピクセルのフロントエンドノイズにより、実際のイベント検出時間は、投影ピクセルごとに異なる。仮に全ての投影ピクセルでの実際のイベント検出の分布をプロットすると、およそ図3に示すような曲線を描く。

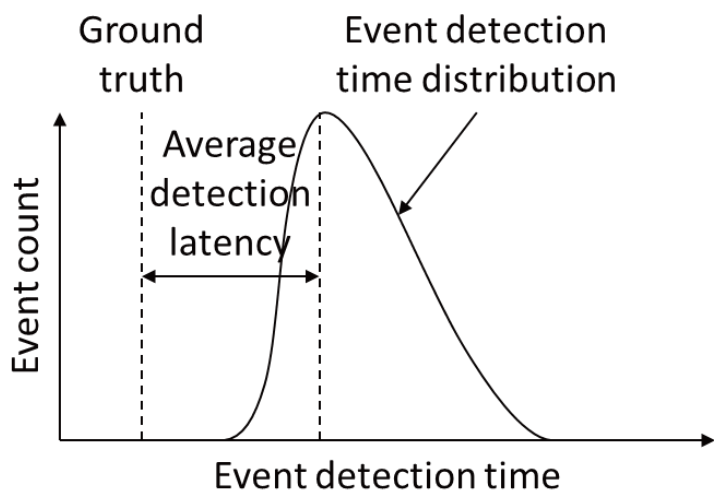


図3) イベント検出時間分布

Ground truth	グラウンドトゥルース
Event detection time distribution	イベント検出時間分布
Average detection latency	平均検出レイテンシ
Event count	イベント回数
Event detection time	イベント検出時間

このイベント検出時間分布は、個々のイベントが高時間分解能(例、 $1\mu\text{s}$ あるいは 1ns)で生成されると同時にタイムスタンプされる場合には、非常に正確に捉えることができる。しかし、そのような正確なタイムスタンプはコストが高い。タイムスタンプ回路を全てのピクセル内部に備えなければならず、ピクセルの複雑さとサイズが大幅に増大することになるからだ。また、イベント検出時間分布は、ピクセルフロントエンドノイズと平均検出レイテンシに関する統計的情報を提供するが、イベント検出の発生順は、グラウンドトゥルースとの相関がまったくない。

実際のところは、そのような正確なタイムスタンプが、現実の使用事例すべてに必要なかは明らかではない。

したがって、多くのニューロモルフィックイベントベースセンサーでは、ピクセルアレイの外部で読み出し時タイムスタンプスキームを適用している。イベントは、ピクセルアレイから読み出されるときに、ピクセルごとに、あるいはグループ(列、行、またはアレイ全体)ごとに、タイムスタンプされる。読み出し時のタイムスタンプには、読み出し時のレイテンシを起因とするエラーイベントのタイムスタンプも含まれる。イベントの読み出しレイテンシは、最大転送速度と瞬間イベント検出率によって変わるので、100ns未満から1msを超えるまでの幅がありうる。結果として、読み出し時タイムスタンプスキームを経由した最終イベントタイムスタンプは、図4のようになる。たとえタイムスタンプが1 μ s、あるいは1nsの単位になっても、タイムスタンプの分布は、瞬間イベント検出率が最大読み出し速度より高ければピクセルフロントエンドノイズについて信頼できる統計的情報を提供できず、タイムスタンプの発生順もグラウンドトゥルースとはまったく相関がない。

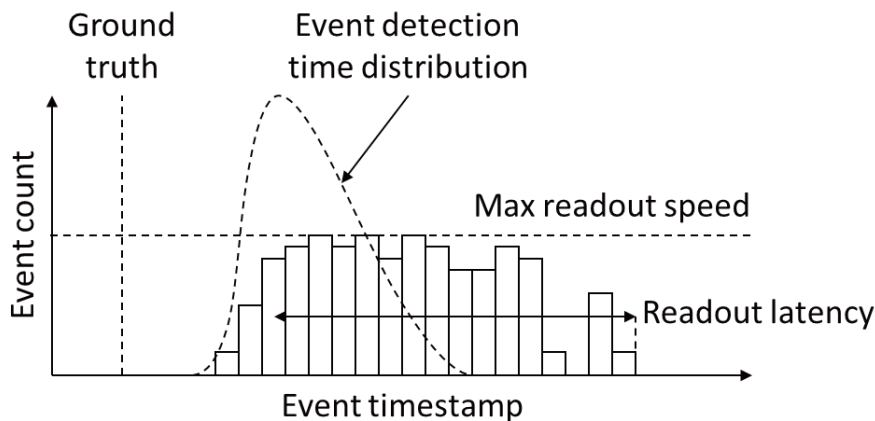


図4) 読み出し時タイムスタンプを使用するイベントタイムスタンプの分布

Ground truth	グラウンドトゥルース
Event detection time distribution	イベント検出時間分布
Event count	イベント回数
Max readout speed	最大読み出し速度
Readout latency	読み出しレイテンシ
Event timestamp	イベントタイムスタンプ

まとめると、ピクセルフロントエンドノイズが原因で、そして最もよく適用されている読み出し時タイムスタンプスキームが原因で、 $1\mu\text{s}$ 、さらに 1ns という小さなタイムスタンプ単位でも、 $100\mu\text{s}$ かさらに 1ms オーダーという粗いタイムスタンプ単位に比べ、グラウンドトゥールースに関して追加で情報が得られることはほとんどない(使用事例による)。逆に、不必要に高い時間分解能があることで、タイムスタンプシステムがいっそう複雑になり、情報の少ない追加データを扱うための通信費と処理コストが上がることになる。いたずらに高い時間分解能は、システムレベルでの費用がかさみ、実証された効果もない。実際に、読み出し時のタイムスタンプスキームにおけるイベントタイムスタンプの有用性も、最大読み出し速度に大きく依存する。イベントの読み出しが速くできるほど、タイムスタンプのエラーも小さくなる。これについては後述する。

これらを考慮すると、引用された低い時間分解能だが高い最大読み出し速度のセンサー(例えば我々のカメラで、時間分解能 $200\mu\text{s}$ 、最大イベントスループット 165MEPS のDVXplorer)は、高い時間分解能だが非常に低い最大読み出し速度のセンサー(例えば我々のカメラで、時間分解能 $1\mu\text{s}$ 、最大イベントスループット 12MEPS のDAVIS346)よりも、実際にははるかに実用的なことがある。

非同期イベント読み出しスキームと同期イベント読み出しスキームの長所と短所は何か？

非同期イベント読み出しスキームとは、イベントがピクセルアレイからホスト受信機へ、クロック制御されていないハンドシェイク通信プロトコルを介して通信が行われることを指す。先のイベントの通信が行われている間に複数の新しいイベントが検出されると、それらは同時発生していると思なされる。これらの同時発生イベントは、仲裁システムにより決定された順に通信が行われる。また、同時発生イベントは、通常読み出し時にタイムスタンプされるので、同じタイムスタンプを共有することはない。非同期イベント読み出しスキームの主な長所は、低い消費電力と短い読み出しレイテンシである。どちらもイベント検出率が最大読み出し速度よりも低い場合に限られた利点である。非同期イベント読み出しスキームの最も大きな短所は、その制限される最大読み出し速度である。

非同期イベント読み出しスキームでは、イベントの通信はイベントの検出により起動される。イベント検出率が低い場合には、読み出しシステムがあまりアクティブではない状態になり、時には待機状態にまでなるので、消費電力は非常に小さくなる。さらに、イベント検出率が最大読み出し速度よりも低い場合には、新たに検出されたイベントの通信は、最小遅延での検出によって直ちに起動される。

一例として、非同期イベント読み出しスキーム付きの我々のカメラDAVIS346は、1イベント通信につき約80nsを必要とする。したがって、イベント検出率が80nsにつき1イベント（およそ12MEPS）より低くなると、DAVIS346の最良事例での読み出しレイテンシは約80nsとなる。

しかし、イベント検出率が増大するに伴い、非同期イベント読み出しスキームは、電力効率が下がる。これは、ハンドシェイク通信プロトコルが、単一イベントの通信のためにピクセルアレイとホスト受信機の間で複数の情報交換を要求するからである。さらに、瞬間的なイベント検出率が最大読み出し速度よりも高くなると、通信の順番待ちをしている同時発生イベントの数が累積し始め、その結果として、読み出しレイテンシが増大する。この読み出しレイテンシの増大は、順番待ちシーケンスを決定する仲裁システムを使用するため、個々のイベントに対して非決定的である。我々のカメラDAVIS346（最大イベントスループット12MEPS）を例にとると、入力刺激により10kの同時発生イベント（全0.1Mピクセル中、約10%）が生じる場合、これらのイベントに対して最悪ケースの読み出しレイテンシは約1msであり、タイムスタンプの大きなエラーを引き起こす。

図5に示すように、新たなイベントが1ms以内で検出された場合、それらのイベントは、仲裁システムによって通信待ちの列に挿入され、読み出しレイテンシと先に累積された同時発生イベントのタイムスタンプのエラーとをさらに増大する可能性がある。

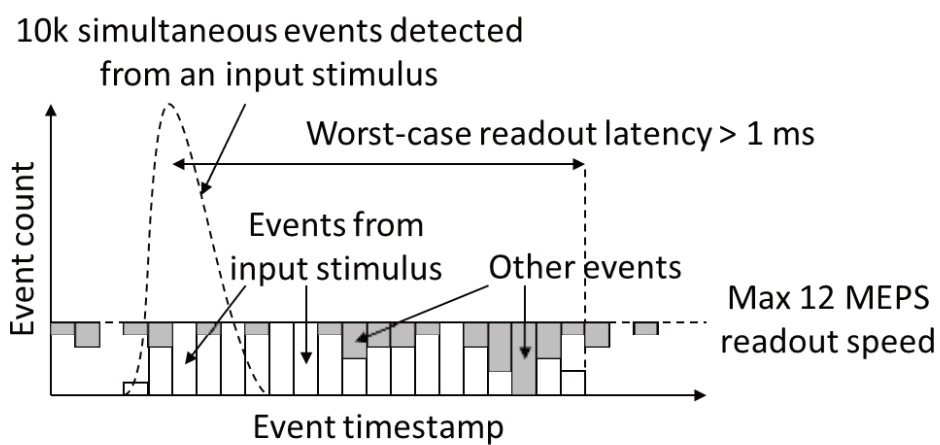


図5) 非同期イベント読み出しスキームにおける最悪ケースの読み出しレイテンシ (DAVIS346)

"10k simultaneous events detected from and input stimulus	1つの入力刺激から検出された10k同時発生イベント
Worst-case readout latency > 1ms	最悪ケースでの読み出しレイテンシ > 1ms
Event count	イベントカウント
Events from input stimulus	入力刺激からのイベント
Other events	他のイベント
Max 12 MEPS readout speed	最大12MEPS読み出し速度
Event timestamp	イベントタイムスタンプ

同期イベント読み出しスキームとは、イベントがピクセルアレイからホスト受信機へとクロック制御のシステムティックスキームを介して通信されることを意味する。同期イベント読み出しスキームは一般的に、グローバルシャッターのようなグローバルイベントサンプリングメカニズムと組み合わせて使用される。それは、ピクセルアレイ全体が既知の(かつ調整可能な)時間間隔で同時にサンプリングされることを意味する。この時間間隔で複数のイベントが検出され、サンプリングされると、これらは同時発生的と捉えられ、上から下へ、左から右へというようなピクセルアレイの位置に従って、システムティックな順番で通信される。同時発生イベントは同じタイムスタンプを共用し、一つのイベントグループとして参照することができる。同期イベント読み出しスキームの主な長所は、高いイベント検出率での低く一定した読み出しレイテンシと、より高速の最大読み出し速度を可能にすることである。同期イベント読み出しスキームの短所は、低いイベント検出率での一定の静的消費電力とイベント読み出しレイテンシである。グローバルイベントサンプリングの結果としての粗いタイムスタンプは、時間分解能への質問に対する回答として前述したように短所ではない。

同期イベント読み出しスキームにおいて、イベントの通信は、一般的には特定のサンプリング率での全てのグローバルイベントサンプリング後にピクセルアレイ外部にある読み出しコントローラにより指示される。そのため、読み出しコントローラは、個々のグローバルサンプリング間隔中に検出されたイベントの数を知らず、個々のグローバルサンプリングに続いて、検出された可能性のあるイベントの通信を起動することになる。これらの定期的サンプリングと通信制御は、例えゼロイベント検出率であっても一定の非ゼロ静的電力消費を引き出す。幸い、設計最適化により、静的電力抛出の大きさは通常、センサー全体の総消費電力と比較しても無視できるほどである。イベント検出率が増大するにつれ、この同期イベント読み出しスキームは電力効率的となる。なぜなら、個々の単一イベントの通信は単にクロック制御された一つの情報交換を求めるからである。

また、読み出しコントローラにより起動される定期的な通信は、検出されたイベントでの最悪ケースの読み出しレイテンシが、低いイベント検出率でも少なくともグローバルイベントサンプリング間隔と同じになることを意味する。

最大読み出し速度よりイベント検出率が低いときは、この最大レイテンシは変わらない。ただし、同期イベント読み出しスキームは、最大読み出し速度が大きいため、イベント検出率が高くても、一定の最悪ケースの読み出しレイテンシを維持できる。例えば、我々のDVXplorerカメラは、デフォルトのグローバルイベントサンプリング率として5kHzの同期イベント読み出しスキームを導入している。それは、低いイベント検出率で、少なくとも200 μ sの最大(最悪ケースの読み出し)レイテンシを意味する。しかし、165MEPSの最大読み出し速度では、DVXplorerカメラは、30k以上の同時発生イベントを処理しながら、400 μ s以下の最大(最悪ケースの読み出し)レイテンシを維持する能力がある(図6に示すように、イベントがグローバルイベントサンプリング間隔の初めに検出され、同じサンプリング間隔内で読み出しが継続し完了したと仮定して)。

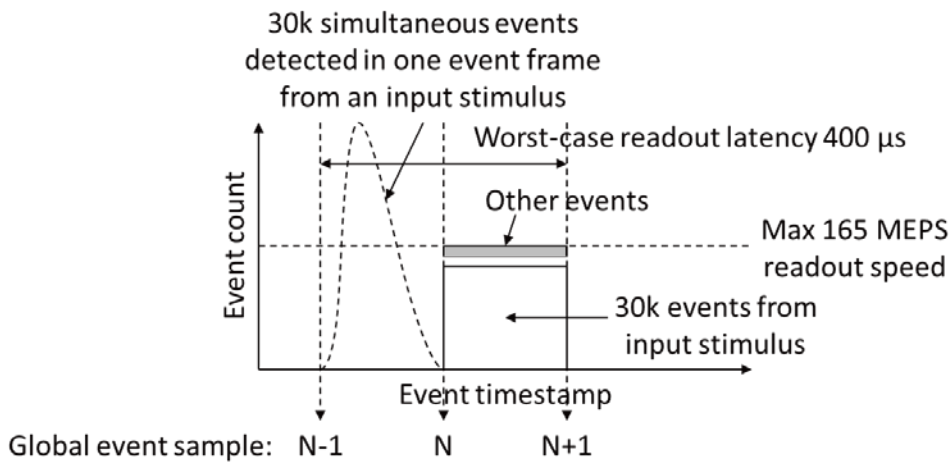


図6) 同期イベント読み出しスキームにおける最悪ケースの読み出しレイテンシ (DVXplorer)

30k simultaneous events detected in one event frame from an input stimulus	入力刺激からの一つのイベント枠で検出された30kの同時発生イベント
Worst-case readout latency 400 μ s	最悪ケースの読み出しレイテンシ 400 μ s
Event count	イベントカウント
Other events	他のイベント
Max 165MEPS readout speed	最大165MEPS読み出し速度
30k events from input stimulus	入力刺激からの30kイベント
Event timestamp	イベントタイムスタンプ
Global event sample:	グローバルイベントサンプル

どのイベント読み出しスキームが優れているか？

その答えは、センサーの分解能と使用事例による。非同期イベント読み出しスキームは、低い分解能のセンサー（例えば0.1MピクセルのDAVIS346センサー）に、あるいはイベント検出率が低いと予想されるような使用事例に、より適切だと言える。同期イベント読み出しスキームは、高い分解能のセンサー（例えば0.3ピクセルのDVXplorerセンサー）、あるいは高いイベント検出率が期待される使用事例に、より適している。使用されているのが非同期か、あるいは同期読み出しスキームかに関係なく、イベント出力データはアドレスイベント表現(AER)フォーマットと同じである。

レイテンシはどのようにして決まるのか？

イベントベースセンサーの**総レイテンシ**は、あるシーンでのグラウンドトゥールースの入力刺激が発生した瞬間から、グラウンドトゥールースデータ入力刺激により生成された対応イベントが読みだされる瞬間までの経過時間として定義される。

総レイテンシは検出レイテンシと読み出しレイテンシで構成される。

● 検出レイテンシ

ピクセルが、対象とするグラウンドトゥールースのデータ入力刺激が発生する瞬間から、結果としてイベントが検出される瞬間まで測定される。

● 読み出しレイテンシ

ピクセル内でイベントが検出された瞬間からイベントが読み出される瞬間まで測定される。

ピクセルフロントエンドノイズにより、検出レイテンシは時折変化し、ピクセル全体に及ぶ。したがって、センサーの検出レイテンシは通常、サンプリングされたピクセルグループ(例を図3に示す)から派生した平均検出レイテンシで参照される。検出レイテンシはシーンの照度に依存し、ピクセルフロントエンド回路バイアス構成にも依存する。シーンの照度が強いほど、ピクセルフロントエンド回路は早く応答する。したがって、検出レイテンシも短くなる。また、ある使用事例の主要要求項目が低い検出レイテンシである場合、ピクセルフロントエンド回路バイアスは、増加するピクセルフロントエンドノイズを犠牲にして反応速度を大きくするよう構成することができる。

読み出しレイテンシは、ピクセルアレイ内のピクセル位置、全体イベント検出率(あるいは同時発生イベントの数)、イベント読み出しスキームに応じて変化する。

信頼性が求められる多くの使用事例では、最小(最良ケースでの読み出し)レイテンシよりもむしろ最大(最悪ケースでの読み出し)レイテンシを知るほうがより重要である。

非同期イベント読み出しスキームを使用したセンサーの総レイテンシの測定結果については、発表された多くの文献で、サンプルグループの平均として総レイテンシを導出するために、ごく一部のピクセルグループのみが刺激され、測定されている。したがって、これらの結果は、読み出しレイテンシが最大レイテンシに近い場合、総レイテンシ性能を反映していない。読み出し速度が低いため、非同期イベント読み出しスキームの最大(最悪ケースの読み出し)レイテンシは実際には比較的に高い。DAVIS346カメラを例にとると、使用事例がシーンの入力刺激として同時発生イベントをピクセルの10%まで生成することを期待している場合、これらイベントの最大レイテンシは約1ms、あるいは、新しいイベントが1msの時間内で検出された場合にはさらに高くなる(前述の図5で説明)。反対に、同期イベント読み出しスキームは読み出し速度が高いため、最大レイテンシは通常、非同期イベント読み出しスキームよりも非常に小さい。例えば、同じ使用事例がDVXplorerカメラに適用され、同時発生イベントをピクセルの10%で生成した場合、最大レイテンシは400 μ sとなる(前述の図6で説明)。

ダイナミックレンジと低照度性能はどのようにして決定されるのか？

センサーのダイナミックレンジとは、センサー作動条件下での最大と最小の照明レベルの比率である。ニューロモルフィック/イベントベース視覚センサーはダイナミックレンジの点で固有の強みを有している。なぜなら、それは対数的な照度の変化を検出し、よって広範囲の照明レベルを包含できるからである。すべてのイメージデバイスと同様に、照明が強くなればなるほど、極度に高い照明レベルがピクセルのフロントエンド回路の対数的変換を飽和するまで(これにより、時間コントラスト感度が低減する)、イベント出力のノイズは小さくなる。極度に低い照明レベルでは、ピクセルフロントエンドの暗信号により信号が弱められ、時間コントラスト感度を低下させる。加えて、極度に低い照明レベルは極度のピクセルフロントエンドノイズの原因となる。それはさらに、実際の信号を劣化させる。

現実世界での最大の自然照明レベルは約100kluxの直射日光からのものである。多くのニューロモルフィック/イベントベース視覚センサーの設計において、100kluxはピクセルフロントエンド回路の対数的変換の飽和レベルよりもはるかに下である。したがって、多くのニューロモルフィック/イベントベース視覚センサーの最大機能的照明レベルは100kluxであると引用されている。100kluxの機能照明レベルよりも高いものを要求することは可能であるが、そのような照明レベルは自然のシーンでは決して到達することはないので無意味である。結果として、ニューロモルフィック/イベントベース視覚センサーのダイナミックレンジは一般的に、より直接的に最低機能的照明レベルの特性により決定される。

最も低い自然照明レベルは完全な闇夜で得ることができる。特定の照明レベル以下では、さらに照明レベルが低下すると、入力刺激により生成される信号イベントからなる出力イベントは少なくなり、出力イベントの多くはノイズイベントである。以前の文献では(本文が書かれた時期で)、センサーが機能していると思なされるには、どれだけの信号とノイズイベントがイベント出力中に必要であるかについて、合意された基準は存在しない。一部で発表された最も低い機能照明レベルの結果は主観的な視覚での印象に基づいている。他に発表された結果では測定基準を指定していない。DVXplorerシリーズセンサーでは、センサーが低照度下でも機能していると思なした控えめで定量的な基準を適用した。それは例えば、センサーが依然機能していると言えるには、ピクセルの99.9%が最小コントラスト感度レベルに等しいコントラストでの入力を検出しなければならない。