Fast-Locking Burst-Mode Clock and Data Recovery for Parallel VCSEL-Based Optical Link Receivers

ABDULLAH IBN ABBAS AND GLENN E. R. COWAN, (Member, IEEE)
Department of Electrical and Computer Engineering, Concordia University, Montreal, QC H3G1M8, Canada

Corresponding author: Abdullah Ibn Abbas (e-mail: a_ibnabb@encs.concordia.ca).

This work was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC) through the Strategic Project Grant Program and in part by the Faculty of Engineering and Computer Science at Concordia University.

ABSTRACT A burst-mode clock-and-data-recovery (CDR) system for a multi-channel vertical-cavity surface-emitting laser (VCSEL)-based non-return-to-zero (NRZ) optical link’s quarter-rate receivers is presented, that utilizes proxy timing recovery for fast turn-on time. The proxy timing recovery scheme takes advantage of correlated data jitter over parallel optical lanes typically deployed in a data-center. Rapid timing recovery of the burst-mode channel is enabled by incrementing/decrementing its phase rotator (PR) control code during idle periods using phase updates from an always-active channel in the link. This work also presents circuit-design techniques to reduce power dissipation during idle times while still enabling fast turn-on time. Simulated in 65 nm CMOS technology, the proposed CDR consumes only 19.5 mW per channel while operating at 10 Gbps/ch and 0.58 mW during its idle-state. Simulation results are presented for the turn-on time with the proposed technique and compared against the turn-on time of a conventional receiver. The proposed technique allows the CDR to lock within 26 unit intervals (ULIs) from when it is powered on irrespective of a 1000 ppm frequency offset between the incoming data and the CDR's reference clock. The complete CDR of each channel occupies an area of 0.045 mm². The proposed scheme introduces only 1.3% of area and 2.6% of on-state power overhead while reducing idle-time power dissipation by 97%.

INDEX TERMS Burst-mode, clock and data recovery (CDR), energy efficiency, injection locked oscillator (ILO), optical link, power-proportional, transimpedance amplifier (TIA).

I. INTRODUCTION
The ever-increasing demand for data rate in data-centers drives the need for high-speed interconnects [1]. Consequently, electrical interconnects are being replaced by optical-fiber interconnects. The main advantage of an optical cable over its electrical counterpart is its reduced frequency-dependent loss. This leads to simpler channel equalization and reduced power dissipation.

Data-centers are using parallel optical links with progressively increasing per-lane data rate to meet overall throughput demands. However, the peak data rate is not required by all links all the time [2]. Indeed, links in data-centers are idle up to 90% of the time [3], but non-useful data packets are still being sent to maintain synchronization. Useful data packets are transmitted intermittently in the form of bursts. However, links dissipate full power even if no useful data are being transmitted. This is graphically shown in Fig. 1 (a), where the flat grey area represents a constant power dissipation level and the hatched area represents the intervals in which useful data are transmitted.

Present-day data-centers contain thousands of servers and their interconnection network consumes around 26% of the total data-center power [4]. As a result, there is a need to reduce the power dissipation of these interconnects when they are idle and reduce the associated cost of cooling data-centers. A promising way to reduce power dissipation is to employ a power-proportional burst-mode optical link, where the link is powered off during idle times and powered on with each burst of data [5]. The ideal scenario for the power dissipation of a power-proportional burst-mode link is shown in Fig. 1 (b) where power is dissipated only when useful data are being transmitted. The main challenges to make such a link attractive for use in data-centers are the reduction of off-state power and the turn-on time.
If a power-proportional burst-mode optical link is put into a sleep mode between data bursts, the receiver clock loses synchronization with the data. When a link is turned on as required by a burst of data, its clock will start at an arbitrary phase with respect to the data and take considerable time to lock to the appropriate phase for low bit-error ratio (BER) sampling (~hundreds of unit intervals (UIs)) [6].

A conceptual transient at the output of a burst-mode receiver is shown in Fig. 2(a). Turning on a burst-mode receiver involves two tasks. First, the average value of the input signal must be acquired for use as a reference level for the receiver’s decision circuit(s). This process is referred to as dc-level recovery. Second, the receiver’s sampling clock must be aligned to capture the data through clock-and-data-recovery (CDR), or in this context, timing recovery. Both of these processes are lengthened if the receiver is turned off or put into an idle-state when no input data are received. Lower power during idle states is usually associated with longer turn-on time. Hence, there is a trade-off between the off-state power and the turn-on time, as illustrated graphically in Fig. 2(b). This work is focused on the timing recovery and its rapid activation from an idle-state.

Recent works on burst-mode solutions for optical networks, wireline electrical, and vertical-cavity surface-emitting laser (VCSEL)-based optical transceivers have been presented in [5]-[11]. In [6], power-hungry limiting amplifiers (LAs) are replaced with a variable-gain amplifier (VGA) to improve the energy efficiency during the active state. The work uses complicated control loops to correct the dc-offset current at the input of the TIA, and an exhaustive CDR search algorithm to reduce the overall turn-on time of the receiver which remains on the order of hundreds of UIs. The work in [5] reduced the duration of timing recovery by using a complex CDR functionality along with a link protocol. Both the settling of the dc-level and the CDR locking take place simultaneously in [8] to reduce the overall turn-on time. The CDR of [11] uses an area intensive architecture that sweeps the oscillator phase after each idle period to achieve phase match with the data signal for the lock, resulting in a long recovery time. However, all of the solutions discussed above address the turn-on time for a single-channel burst-mode receiver. These approaches forgo opportunities arising from the parallel nature of many VCSEL-based links. This work leverages correlation in the timing among parallel channels along with a commonly used CDR topology to achieve the fastest CDR locking. Without additional complex circuitry as compared to the other published works it has the minimum area and power overhead. Since the design is mostly digital, its area and power are also attractive, and will improve with technology scaling.

The proposed work presents an energy-efficient burst-mode CDR for multi-channel parallel links for fast CDR lock time, as initially laid out in [12]. A typical phase interpolator-based CDR [13], as shown in Fig. 3, is used. It consists of a clock synthesizing block and a peripheral loop. The main circuit blocks of the peripheral loop are a phase detector, a loop filter, phase-rotator (PR) control logic, and a PR.

The work has two features: 1) low-power operation during the idle state is enabled by turning off all circuits in the CDR except the PR control logic block that increments/decrements the PR’s control code; 2) a fast CDR lock time is achieved by updating the PR control code (in the PR control logic) during idle times using a phase update signal from an active parallel channel.

The remainder of the paper is organized as follows: Section II briefly highlights the background of the parallel link’s properties, and Section III describes the architecture, implementation details, analysis of jitter tolerance and tolerable frequency offset of the proposed CDR. Simulation results for the fast turn-on time of the proposed technique are presented in Section IV. Section V concludes the work and includes a comparison with other published work.

II. BACKGROUND

To meet high aggregate throughput requirements in datacenters, parallel links are widely used [14], [15]. Each parallel link has a near-identical channel and transceiver. Transmitted data streams over parallel links are synchronous from lane to lane as they share the same transmitter-side clocking circuitry and a common transmitter-side reference clock. A possible mismatch exists in the path length among these channels. However, in serial links, the mismatch in length is not a problem in clock recovery since each lane has its own CDR.

Fig. 4 shows the phase offset between the data at the receivers of two parallel channels relative to the receiver’s reference clock in the presence of parts per million (ppm) frequency offset between the data streams and the receiver’s reference clock. The data streams have no ppm offset from one lane to another. One complete cycle around the circle
Fig. 2. (a) Graphical representation of transients in a burst-mode link and (b) conceptual illustration of response time as a function of off-state power.

Fig. 3. A typical phase-interpolator-based CDR.

Fig. 4. Exemplary phase relationship between the data of two channels in the presence of a ppm frequency offset between data streams and the receiver’s reference clock.

Fig. 5. (a) Phase relationship of two data streams, their recovered clocks and the receiver’s reference clock, shown at the end of a burst and the start of the next burst on the burst-mode channel. Shown when the burst-mode channel’s PR code is not updated. (b) The aforementioned phase relationship with the proposed proxy timing recovery scheme where the PR code of the burst-mode channel is updated between bursts.

represents 2π or one reference clock cycle. \( M \) denotes the number of data UIs to cover one complete cycle, which is inversely proportional to the ppm frequency offset. Due to the ppm offset, the phase offset for data on channels 1 and 2 increases between \( t = 0 \) and \( t = N \) UI as well as during successive steps of \( N \) UI. However, the phase offset between channels 1 and 2 remains small due to their shared Tx clock. In this figure, \( N \) is chosen arbitrarily to be \( M/5 \), giving 5 N-UI steps for the phase offset of the data channels to advance around the circle. The small phase offset between channels 1 and 2 will be exploited in a fast recovery of the clock of a burst-mode receiver when it turns on.

The phase relationship of the data streams and their recovered clocks of an always-on channel and a burst-mode channel with respect to the receiver’s reference clock is shown in Fig. 5 (a). \( \phi_1 \) is the phase difference between the receiver’s reference clock and the recovered clock of the always-on channel at \( t=0 \). \( \phi_2 \) is the phase difference between the receiver’s reference clock and the recovered clock of the burst-mode channel at \( t=0 \). In this case, the PR control code of the burst-mode channel is not updated while the channel is idle. The right-hand side of Fig. 5 (a) shows the phase relationships after an idle period of \( M/2 \). The phase of the always-on channel’s data and recovered clock has advanced by \( \pi \), reaching \( \phi_1 + \pi \). Since the burst-mode channel’s PR control code was not updated, its phase remains at \( \phi_2 \), but is misaligned to the burst-mode channel’s data which has also advanced by \( \pi \) during the idle period. Using the burst-mode
channel’s clock to sample the next burst will lead to a severe degradation in bit-error ratio. Allowing the CDR to lock using conventional loop dynamics will incur a time delay of possibly hundreds of UIs.

The phase relationship is repeated in Fig. 5 (b) where the PR code of the burst-mode channel is updated between bursts using the proposed timing recovery scheme. Even though most of the burst-mode channel’s CDR has been powered down between t = 0 and t = M/2 UI, its PR code has been updated using duplicate PR updates from the always-on channel. Therefore, it powers up with a phase offset suitable to capture data from the burst-mode channel.

The advantage of the proposed scheme is that the clock of the burst-mode receiver remains aligned with the data stream from the very first stable clock cycle. The phase detector output of the always-on channel is used as a proxy for phase updates during idle intervals. Hence, this method of timing recovery is referred to as “proxy” timing recovery. The “collaborative” timing recovery scheme used in [16] recovers a global clock from the simultaneous participation of all the channels in the link and is different from that proposed here.

III. PROPOSED ARCHITECTURE AND DESIGN

Twelve or more parallel optical-fiber channels are frequently deployed in data-centers [17] where each channel has an independent CDR at the receiver side. A conceptual block diagram of the proposed twelve-channel burst-mode link is shown in Fig. 6 which without the connection among CDRs is similar to a conventional architecture presently deployed in data-centers. The proposed architecture is compatible with the approach in [17] so long as it is assumed that even during periods of low link activity, at least one channel is active, maintaining synchronization. The remaining channels are turned on or off independently depending on the per-channel workload. Each channel has two power states: i) idle-state and ii) on-state. An external half-rate reference clock (5 GHz) is provided for all the CDRs on the receiver side. During the idle-state, each burst-mode receiver receives a 1/4th rate clock from the always-on channel to detect the beginning of a burst of data, and its PR control code is incremented/decremented using phase updates from the always-on channel. Since the preamble of a burst uses 1/5th rate data, the 1/4th rate sampling clock can have arbitrary phase. Although the always-on channel can take any arbitrary location, to maximize the jitter correlation between the recovered clock on the always-on channel and the data on the burst-mode channels, the always-on channel is located in the middle [16].

To investigate opportunities to reduce the turn-on time in parallel optical links, a three-channel receiver prototype is designed, as shown in Fig. 7. It consists of one always-on channel and two burst-mode channels; the latter operate with rapid on/off functionality to improve energy efficiency during idle times. All channels can operate at 10 Gbps. In the idle-state, most of the circuitry of the burst-mode channel’s CDR is turned off to reduce power dissipation. Also, the envisioned front-end of the receiver is configured to have reduced power dissipation and bandwidth but equal gain during the idle-state. The proposed link protocol and the circuit details of the building blocks for the proxy timing-recovery CDR are described in the following subsections.

A. PROPOSED BURST-MODE LINK PROTOCOL

The proposed link protocol is shown in Fig. 8 (a). When configured in the idle-state, the front-end is a fully operational receiver, with reduced bandwidth and power dissipation compared to when it operates in the on-state. The

Fig. 6. Conceptual block diagram of the proposed burst-mode receiver for a twelve-channel parallel optical link showing one always-on channel and eleven burst-mode channels.
sensing of a data burst is marked by the detection of a few positive edges in an incoming 1/5\textsuperscript{th} rate bit stream. This is achieved by a sensing circuit through a continuous sampling process by a 1/4\textsuperscript{th} rate clock available from the always-on channel. To power on a receiver from its idle-state, an alternating sequence of five “1s” and five “0s” (ON-BITS) starts the preamble sequence. These ON-BITS, captured by a 1/4\textsuperscript{th} rate decision circuit, produce positive edges that switch the receiver from idle-state to on-state. In this prototype, three positive edges confirm a data burst, thus avoiding any false glitches arising from transients that might appear when the receiver is switched from its on-state to idle-state.

The sequence of 5 bits for each “1” and “0” is to guarantee the three positive edges in the sampled data with a 1/4\textsuperscript{th} rate clock available from the always-on channel. The 1/4\textsuperscript{th} rate sampling clock from the always-on channel is aligned with the midpoint of the data on the always-on channel, but the incoming data on the burst-mode channel has an arbitrary phase deskew with the sampling clock of the always-on channel. This phase deskew is due to the non-phase-matched clocking architecture used for the distribution of the recovered clock from the always-on channel (Fig. 7), and due to a mismatch in the path lengths of the channels, shown in Fig. 8 (b). The figure shows possible incoming data phases (i-iii) with respect to the clock from the always-on channel. This clock is used only to detect the incoming burst and is not used for sampling the rest of the burst of data on the burst-mode channel once it is activated. The corresponding sampled bits are shown in Fig. 8 (c), and the curved arrows denote positive transitions. The first “0” in each row of sampled data represents the last sampled bit during the idle-state. Even if the first clock edge falls on the metastable position of the data, the circuit with the help of the proposed sequence is able to detect three positive edges. All possible combinations for this case are also presented (square box).

The rest of the preamble sequence (DC-PREAMBLE) is a repeated “1100” pattern of 24 UIs during which the output dc-level is recovered. The useful data follow the preamble sequence, and the link is fully operational. Since the PR control code was updated using phase updates from the always-on channel, the CDR in the burst-mode channel starts from the correct phase with respect to its incoming data burst.

At the end of a data burst, receiving 20 consecutive 0s brings the receiver into the idle-state.

**B. SENSE CIRCUIT DESCRIPTION**

![Diagram](image)

The sense circuit (SENSE CKT), shown in Fig. 9, senses the incoming burst of data in 20 UIs and generates the ON-OFF signal to toggle the receiver between the idle-state and the on-state. When ON-OFF = 0, the receiver is in the idle-state. The recovered clock from the always-on channel (CLK1) passes to the latch L1 which samples the output of the front-end in anticipation of an incoming burst of data, ultimately producing three rising edges. The DFFs are clocked by the output of the latch. With the three rising edges, the outputs of the three DFFs are 1, resulting in ON-OFF=1, and thus, the clock to latch L1 is gated. Following this, the receiver enters the on-state. When the receiver needs to be placed in the idle-state, the “Off Detect” block (clocked by its own recovered clock, CLK2) generates a reset pulse (RST), after sensing twenty consecutive “0s”, which resets the three DFFs inside the “Burst Detect” block and turns ON-OFF=0.

The generation of the RST pulse is governed by the following equation where signals A, B, and Y are shown in Fig. 9:

\[ Y = (A \oplus B) \cdot B \]

\[ = \bar{A} \cdot B. \]  

The “Off Detect” circuit in the proposed work has been optimized based on the available four data sampling latches for a quarter rate receiver.

**C. CDR ARCHITECTURE AND DESCRIPTION**

A phase interpolator (PI)-based CDR [13] is used in this work. The CDRs of the always-on channel and one burst-mode channel are shown in Fig. 10. The only difference in the CDR of the burst-mode channel is the selection circuits.
for the UP/DN and TRIG signals denoted by the shaded region. CLK1 and CLK2 are the clocks generated by the CDRs of the always-on channel and the burst-mode channel, respectively. Early and Late signals are generated from data (D1-D4) and edge (E1-E4) samples by the Alexander phase detector (PD) and majority voting circuits. They are then processed by the loop filter to create an UP/DN signal (1=UP, 0=DN) and a triggering pulse (TRIG) for the PR control logic. The PR control logic produces a 64-bit code for the PR. The PR generates a differential clock signal, whose rotation covers the whole clock period, and drives the “8-phase generator” block. The outputs of the “8-phase generator” then serve as the clock for the data and edge sampling latches.

The two channels of the proposed architecture interact with each other in such a way that the UP/DN and TRIG signals from the always-on channel feed the PR control logic of the burst-mode channel through the selection multiplexers when the ON-OFF signal is low (idle-state). Also, the generated clock (CLK1) from the CDR of the always-on channel is
supplied to the SENSE CKT of the burst-mode channel through a gating circuit. In the idle-state, the SENSE CKT continuously checks for the start of a data burst with the help of CLK1.

The circuit details and the functionality of the loop filter, consisting of a 4-bit bidirectional shift-register and a finite state machine (FSM), are shown in Fig. 11. The minimum output update rate of the loop filter is 6 clock cycles. The purpose of using a 4-bit shift register is to have a balance between jitter tolerance and false phase steps due to the overall delay of the CDR loop. CLK2 passes to the loop filter only when Early and Late signals are unequal, giving a gated clock. GCK [18]. The advantage of clock gating the loop filter is to reduce the dynamic power of the CDR. The TRIG pulse is generated only when both Q1 and Q4 are high.

The 2.5 GHz differential in-phase and quadrature clock signals, generated by a single 5 GHz reference clock (REF_CK) with the help of an I/Q generator, drive a 128-step CML PR. The 5 GHz REF_CK would be generated using a shared PLL across all twelve lanes, as is commonly done in multi-lane interfaces [16] and distributed. The proposed scheme does not rely on precise matching of the REF_CK phase at each CDR since each lane has its own PR. The output of the PR goes to an 8-phase generator. The 45º spaced output clock phases then drive the high-speed (data sampling) latches.

The implemented CML PR is shown in Fig. 12 (a) with circuit level details of one cell [19] and total phase steps of 128 in one clock period (32 phase steps per UI). The low complexity architecture has a conventional round interpolation, giving a gated clock. The advantage of clock gating the loop filter is to reduce the dynamic power of the CDR. The TRIG pulse is generated only when both Q1 and Q4 are high.

The implemented CML PR is shown in Fig. 12 (a) with circuit level details of one cell [19] and total phase steps of 128 in one clock period (32 phase steps per UI). The low complexity architecture has a conventional round interpolation, giving a gated clock. The advantage of clock gating the loop filter is to reduce the dynamic power of the CDR. The TRIG pulse is generated only when both Q1 and Q4 are high.

The implemented CML PR is shown in Fig. 12 (a) with circuit level details of one cell [19] and total phase steps of 128 in one clock period (32 phase steps per UI). The low complexity architecture has a conventional round interpolation, giving a gated clock. The advantage of clock gating the loop filter is to reduce the dynamic power of the CDR. The TRIG pulse is generated only when both Q1 and Q4 are high.

The implemented CML PR is shown in Fig. 12 (a) with circuit level details of one cell [19] and total phase steps of 128 in one clock period (32 phase steps per UI). The low complexity architecture has a conventional round interpolation, giving a gated clock. The advantage of clock gating the loop filter is to reduce the dynamic power of the CDR. The TRIG pulse is generated only when both Q1 and Q4 are high.

The implemented CML PR is shown in Fig. 12 (a) with circuit level details of one cell [19] and total phase steps of 128 in one clock period (32 phase steps per UI). The low complexity architecture has a conventional round interpolation, giving a gated clock. The advantage of clock gating the loop filter is to reduce the dynamic power of the CDR. The TRIG pulse is generated only when both Q1 and Q4 are high.

The implemented CML PR is shown in Fig. 12 (a) with circuit level details of one cell [19] and total phase steps of 128 in one clock period (32 phase steps per UI). The low complexity architecture has a conventional round interpolation, giving a gated clock. The advantage of clock gating the loop filter is to reduce the dynamic power of the CDR. The TRIG pulse is generated only when both Q1 and Q4 are high.

The implemented CML PR is shown in Fig. 12 (a) with circuit level details of one cell [19] and total phase steps of 128 in one clock period (32 phase steps per UI). The low complexity architecture has a conventional round interpolation, giving a gated clock. The advantage of clock gating the loop filter is to reduce the dynamic power of the CDR. The TRIG pulse is generated only when both Q1 and Q4 are high.

The implemented CML PR is shown in Fig. 12 (a) with circuit level details of one cell [19] and total phase steps of 128 in one clock period (32 phase steps per UI). The low complexity architecture has a conventional round interpolation, giving a gated clock. The advantage of clock gating the loop filter is to reduce the dynamic power of the CDR. The TRIG pulse is generated only when both Q1 and Q4 are high.

The implemented CML PR is shown in Fig. 12 (a) with circuit level details of one cell [19] and total phase steps of 128 in one clock period (32 phase steps per UI). The low complexity architecture has a conventional round interpolation, giving a gated clock. The advantage of clock gating the loop filter is to reduce the dynamic power of the CDR. The TRIG pulse is generated only when both Q1 and Q4 are high.

The implemented CML PR is shown in Fig. 12 (a) with circuit level details of one cell [19] and total phase steps of 128 in one clock period (32 phase steps per UI). The low complexity architecture has a conventional round interpolation, giving a gated clock. The advantage of clock gating the loop filter is to reduce the dynamic power of the CDR. The TRIG pulse is generated only when both Q1 and Q4 are high.

The implemented CML PR is shown in Fig. 12 (a) with circuit level details of one cell [19] and total phase steps of 128 in one clock period (32 phase steps per UI). The low complexity architecture has a conventional round interpolation, giving a gated clock. The advantage of clock gating the loop filter is to reduce the dynamic power of the CDR. The TRIG pulse is generated only when both Q1 and Q4 are high.

The implemented CML PR is shown in Fig. 12 (a) with circuit level details of one cell [19] and total phase steps of 128 in one clock period (32 phase steps per UI). The low complexity architecture has a conventional round interpolation, giving a gated clock. The advantage of clock gating the loop filter is to reduce the dynamic power of the CDR. The TRIG pulse is generated only when both Q1 and Q4 are high.

The implemented CML PR is shown in Fig. 12 (a) with circuit level details of one cell [19] and total phase steps of 128 in one clock period (32 phase steps per UI). The low complexity architecture has a conventional round interpolation, giving a gated clock. The advantage of clock gating the loop filter is to reduce the dynamic power of the CDR. The TRIG pulse is generated only when both Q1 and Q4 are high.

The implemented CML PR is shown in Fig. 12 (a) with circuit level details of one cell [19] and total phase steps of 128 in one clock period (32 phase steps per UI). The low complexity architecture has a conventional round interpolation, giving a gated clock. The advantage of clock gating the loop filter is to reduce the dynamic power of the CDR. The TRIG pulse is generated only when both Q1 and Q4 are high.

The implemented CML PR is shown in Fig. 12 (a) with circuit level details of one cell [19] and total phase steps of 128 in one clock period (32 phase steps per UI). The low complexity architecture has a conventional round interpolation, giving a gated clock. The advantage of clock gating the loop filter is to reduce the dynamic power of the CDR. The TRIG pulse is generated only when both Q1 and Q4 are high.
Fig. 11. Circuit details of the loop filter and the graphical illustration of the generation of UP/DN and TRIG signals.

Fig. 12. (a) PR circuit with transistor level diagram of one cell. (b) PR control logic generating 64 control bits for the PR. (c) Block diagram of the 8-phase generator incorporating three cascaded ILOs for the generation of eight clock phases with 45° phase spacing.

UP/DN and TRIG signal from the loop filter is shown in one direction only when UP/DN=1. Similarly, for UP/DN=0, the DACs will be updated in the opposite direction. The 8-phase generator consists of three cascaded injection-locked oscillators (ILOs), shown in Fig. 12 (c).

Each ILO is a four-stage, cross-coupled, pseudo-differential ring oscillator with four differential injection points. For the first ILO, the differential output of the PR is received at one differential injection point, and the other three injection points are disabled. The other two ILOs receive the four differential outputs of the preceding ILO as the injection signals for their four differential injection points. The four differential outputs of the third ILO form eight clock phases, for data and edge sampling, with 45° phase separation. The use of three cascaded ILOs provides better phase separation (each with 45°) compared to only two cascaded ILOs.

The CDR of the always-on channel is in operation at all times with full power dissipation. However, during its idle state, the CDR of the burst-mode channel consumes a small fraction of its on-state power dissipation due to activity in the PR control logic, necessary to keep the PR code updated by the always-on channel. All other digital blocks in the CDR remain inactive and dissipate no power. The I/Q generator, the PR, and the 8-phase generator are turned off through the ON-OFF signal and dissipate only leakage power in the idle state. All the switching in the proposed circuit is done simultaneously with the generated ON-OFF signal.
D. CDR LOOP DYNAMICS, STABILITY AND INTER-LANE TRACKING

The CDR used in this design consists of a clock synthesis circuit and a peripheral loop. The clock synthesis circuit includes the external reference clock and the I/Q generator, whereas the peripheral loop comprises the PD, loop-filter, PR control logic, PR, and the ILOs. The CDR has a first-order loop (Fig. 13) and hence it is unconditionally stable. In Fig. 13, the value of $K_o$ is 1.

The phase resolution achieved by the CDR is given by [20]

$$\text{Phase Resolution (}\Delta\Phi) = \frac{\text{Clock Period (}T_{\text{CK}}\text{)}}{\text{No. of Accumulator States}}$$

$$= \frac{4}{128} \text{ UI}$$

$$= 31.25 \text{ mUI}$$

(2)

where 1 UI is the bit period of the incoming data, and the number of accumulator states achieved by two 32-bit shift registers (with a cascading arrangement) in the PR control logic is 128.

The maximum tolerable frequency error ($\Delta f$) between the incoming data and the nominal frequency ($f_{\text{nom}} = 2.5 \text{ GHz}$) of the receiver’s reference clock is

$$F_{\text{tol, max}} = \frac{\Delta f}{f_{\text{nom}}} \approx \frac{\text{Phase Resolution}}{\text{Min. Phase update interval}}$$

$$= \frac{\Delta \Phi}{6 \times T_{\text{CK}}}$$

$$= \frac{31.25 \text{ mUI}}{6 \times (4 \times 1 \text{ UI})} = 1302 \text{ ppm.}$$

(3)

The factor 6 in the above equation represents the minimum update clock cycles for UP/DN signals out of the loop filter.

The jitter tolerance (JTOL), which is defined as the maximum sinusoidal jitter amplitude $A$ on the incoming data that the designed CDR can tolerate for a given jitter frequency $f_j$ is given by

$$A < \frac{\Delta \Phi}{2\pi} \frac{1}{f_j} \frac{f_{\text{nom}}}{6}.$$  

(4)

Clearly, the higher the frequency of jitter, the lower the amplitude of jitter that can be tolerated. Both (3) and (4) give theoretical values exceeding what is seen in a real implementation. A reasonable choice of the design parameters was made in this work, and the JTOL curve resulting from (4) is comparable with published work [21].

The use of phase detection from one channel in the updating of the PR code in another channel raises the question of whether the correlation of data jitter seen on adjacent channels leads to detrimental differential jitter, because of the difference in data path length and REF_CK path length between channels [22]. A delay mismatch due to combined data path and REF_CK path length differences between channels could exceed a few UIs over a 100 m interconnect. Considering clock jitter on the first channel given by [22]

$$J_{C1} = J_A \sin(2\pi f_j t),$$  

(5)

where $J_A$ is the jitter amplitude and $f_j$ is the jitter frequency, the clock jitter on the second channel in the presence of a data and REF_CK path length mismatch is then given by

$$J_{C2} = J_A \sin(2\pi f_j (t + mT_b)),$$  

(6)

where $m$ is the combined delay mismatch in number of UIs and $T_b$ is the bit period. Therefore, the differential jitter between the two clocks is given by

$$J_D = J_{C1} - J_{C2}.$$  

(7)

This differential jitter is irrelevant when both lanes operate with separate CDRs. However, when the PR control code on the burst-mode channel is updated based on the always-on
channel, \( J_C1 \) is imposed on the burst-mode channel's clock whereas its data has \( J_C2 \). If \( J_D \) exceeds the width of the bathtub curve, errors will occur when the burst-mode channel is activated using the always-on channel’s PR updates. For a given jitter amplitude and jitter frequency, as the delay mismatch increases, the differential jitter increases and becomes out of phase, resulting in an increased (BER).

The maximum jitter amplitude tolerated by a single lane, given by (4) introduces a timing error \( J_D \), between the clocks of the always-on and burst-mode channels, less than 0.1 UI when the combined data and the REF_CK path difference is up to 78 UIs. This path length mismatch is unlikely to occur in a link of 100m or less. Therefore, updating the PR code of the burst-mode channel during its idle-state will provide a clock, accurate to within 0.1 UI at the onset of a data burst where jitter is considered to be correlated. The effect of uncorrelated jitter is considered in Section IV.

## IV. SIMULATION RESULTS

![Fig. 15. Theoretical and simulated JTOL of the proposed CDR.](image)

![Fig. 16. Phase delay plot of the phase rotator against phase steps.](image)

![Fig. 17. Mismatch variation at phase rotator code with (a) minimum phase delay error, (b) maximum phase delay error.](image)

![Fig. 18. Process variation at phase rotator code with (a) minimum phase delay error, (b) maximum phase delay error.](image)

![Fig. 19. (a) Phase delay variation as a function of run number with PR codes resulting in phase step number 0 and phase step number 16. (b) Phase delay difference between these two PR codes as a function of run number.](image)

![Fig. 20. Tolerated uncorrelated jitter by the proxy timing recovery plotted against the idle period with three different jitter periods of 1000 ns, 242 ns, and 121 ns.](image)

![Fig. 21. Simulated (extracted) locking behavior of the complete CDR.](image)
The functionality of the prototype of the proposed burst-mode receiver (Fig. 7) is presented in this section. The proposed CDR is designed in a 65 nm, 1-V CMOS process. The total power dissipation of the CDR of each channel operating at 10 Gbps is 19.5 mW (1.95 pJ/bit) during the on-state and reduces to 0.58 mW in the idle-state. The power breakdown of the CDR is presented in Table I. The layout of the entire CDR of one channel is shown in Fig. 14.

The theoretical value of the JTOL of the CDR used in this work is plotted in Fig. 15. Two different jitter frequencies are selected (4.14 MHz and 8.28 MHz) for simulating the performance of the CDR with added sinusoidal jitter at these frequencies. The maximum amplitude of jitter tolerated by the CDR was noted (limiting the phase offset between the clock and the data to within 0.3 UI peak-peak). The noted values of the jitter amplitude are presented in Fig. 15 and are comparable to (4).

There are several non-idealities that stress the performance of the CDR, such as the nonlinearity in the PR, process variation and mismatch effects on the phase interpolator, correlated and uncorrelated jitter. The PR delay with each phase step is plotted in Fig. 16 and the maximum phase offset of 6 ps from an ideal value is obtained for a given PR code.

A. EFFECTS OF PROCESS VARIATION

Local transistor mismatch will lead the phase interpolators used in the CDRs of the always-on channel and the burst-mode channels to behave differently. The effect of mismatch on the phase delay for the PR codes that result in the minimum and the maximum phase delay error (Fig. 16), is shown in Fig. 17 for 100 Monte-Carlo iterations. The phase delays are with respect to the ideal values of delay corresponding to those codes. The means and the standard deviations in the two cases are -2.3 ps and 3.68 ps, and 1.3 ps and 5.4 ps, respectively. The effect of global process variation is shown in Fig. 18 for these PR codes and the corresponding means and standard deviations are 1.4 ps and 6.8 ps, and 2.1 ps and 8.7 ps, respectively. The deviation, for a given PR code, from the average value due to global process variation and local transistor mismatch will be handled by the proxy timing-recovery as static phase offset. However, the effect of local transistor mismatch on the linearity of the PR is of concern. In Fig. 19 (a), the PR delay is plotted against the run number for phase step number 0 and phase step number 16 (50 ps delay between the two different PR codes). The phase variation between these two codes is highly correlated over the iteration number. In Fig. 19 (b) the difference between the delay for PR codes 0 and 16 is plotted, giving a variation in delay of ~3 ps. Therefore, the PRs are robust against local mismatch and the proxy timing recovery scheme achieves lock when the CDR turns on.

B. EFFECTS OF JITTER

Without any data jitter there is a maximum allowable frequency offset between the sampling clock and the data for phase tracking. However, in the presence of a correlated jitter, a lower frequency offset can be tolerated which depends on the amplitude and frequency of the jitter. For an ideal data eye, the maximum allowable phase offset between the sampling clock and the data, resulting from an uncorrelated jitter, is 1 UI_p for a specified BER. In the case of completely uncorrelated jitter, a differential jitter of 1 UI_p (and hence 1 UI_p clock and data phase offset) is due to an uncorrelated jitter amplitude of 0.5 UI_p at any given jitter frequency. This uncorrelated jitter between the always-on channel and the burst-mode channel has an adverse effect only when the burst-mode CDR turns on from the idle-state. During the active state, each CDR can track this jitter and remains locked.

If the idle time is longer than the period of the jitter, then the maximum tolerable amplitude of differential jitter (due to an uncorrelated jitter) is 1 UI_p. However, if the time period of the jitter is longer than four times the idle period, the CDR can sustain uncorrelated jitter with an amplitude that results in a differential jitter of more than 1 UI_p. Under this condition of uncorrelated jitter, the phase offset between the sampling clock and the data remains within 1 UI_p (neglecting other non-idealities). The tolerance for uncorrelated jitter at jitter periods of 1000 ns, 242 ns, and 121 ns against the idle period is plotted in Fig. 20. The longer the period of the jitter compared to the idle period, the higher the amplitude of differential jitter that can be tolerated by the CDR. For example, if the idle period is 40ns, a 1 MHz uncorrelated jitter can be tolerated by the proposed proxy timing-recovery with a differential jitter amplitude (considering anticorrelation) as high as 4.02 UI_p.

C. LOCKING BEHAVIOR

In the proposed design, the overall lock time of the CDR depends mainly on the injection locking of the ILO to the injected clock signal. The extracted simulation (C+CC) of the overall locking behavior of the CDR is presented in Fig. 21, and it takes 1.9 ns to phase lock the third ILO to the reference clock (shown in grey) within a phase error of 15 ps (0.15 UI).

The usefulness of the proxy timing-recovery becomes prominent in recovering the timing information quickly in the presence of a long period without data and with a ppm frequency offset. The transient behavior of the receiver is presented in Fig. 22 and Fig. 23 with a continuous data pattern input to the always-on channel and burst-mode patterns to the burst-mode channels. The simulation starts with a continuous run of pseudorandom binary sequence (PRBS) 2^21-1 on the always-on channel and burst-mode patterns on the burst-mode channels (also PRBS 2^2-1). To take into account the nonlinearity of the PR in the simulation, the input data pattern of the second burst-mode channel is phase shifted by ½ UI, forcing its PR codes to have an offset with the always-on channel. This offset is around 16 phase steps ahead that of the always-on channel, such that when the burst-mode channel turns on, its PR will have near maximum phase delay error (due to the phase update signal from the always-on channel to

2 VOLUME XX, 2017
increment the code by 16 phase steps during the idle period of 40 ns. Simulated plots are presented after an initial period of time over which the sampling clock on the always-on channel and the burst-mode channels are aligned separately to the optimal sampling position. Fig. 22 and Fig. 23 represent the case without the proxy timing-recovery scheme and the case with the proxy timing-recovery, respectively. In the absence of a contribution from the always-on channel to update the control code of the PR in the burst-mode channels, the 1000 ppm frequency offset causes the data phase to drift away
from the reference clock phase between bursts. Thus, when the burst-mode channels turn on following 40 ns without data, their clocks do not lock within the preamble period. A phase difference of ~40 ps is introduced between the sampling clock and the data (Fig. 21). The effect of the PR nonlinearity, as expected, does not have a significant contribution in the phase difference at the end of the idle period. However, in the presence of the phase update contribution from the always-on channel, the PR code is updated, and the sampling clocks of the burst-mode channels have the correct phase at the end of the preamble period of 49 UIs (Fig. 23).

In Fig. 24, the simulation includes correlated jitter in addition to the ppm frequency offset. However, due to the presence of 4.14 MHz correlated jitter with amplitude 0.5 UIp-p, the offset frequency has been reduced to 300 ppm. At the end of the idle period, the clocks are aligned within the preamble period.

Fig. 23. The simulation is repeated with proxy timing recovery scheme and only the data and the sampling clocks of the burst-mode channels are shown because the data and the sampling clock of the always-on channel always remain in lock condition after certain period of initialization. The figure shows locking within the preamble period of 49 UIs in the presence of 1000 ppm frequency offset (burst-mode channel-2 includes the PR nonlinearity).

Fig. 24. The simulation is with proxy timing recovery in the presence of 300 ppm frequency offset (burst-mode channel-2 includes the PR nonlinearity) and 4.14 MHz correlated jitter of amplitude 0.5 UIp-p. The clocks of the burst-mode channels are found to be locked successfully within the preamble period.

Fig. 25. Calculated phase difference (Δt) between the data and the sampling clock resulting due to off period.
The presented ~40 ps difference between the sampling clock and the data is due to an inactive period of only 40 ns chosen for simulation feasibility. A graphical representation of the calculated phase difference (in time, At) between the sampling clock and the data, resulting from different off periods and 1000 ppm frequency offset, is shown in Fig. 25.

Using the proposed technique, irrespective of the length of the off period, the clock remains almost aligned with the data when the CDRs of the burst-mode channels turn on. However, when there is a ppm frequency offset, consecutive identical digits (CIDs) on the always-on channel will result in a phase offset between the sampling clock and the data of the always-on channel and in turn, the burst-mode channel. For a typical ppm frequency offset and a differential jitter of 0.25 UI, the designed CDR can tolerate at most 83 CIDs. In this design, the timing recovery depends mostly on the locking time of the ILO. Therefore, the technique presented here proves to be effective in fast timing recovery in a burst-mode multi-channel application.

V. COMPARISON AND CONCLUSION

This work demonstrates a new approach for fast timing recovery, based on a proxy timing-recovery scheme, and power reduction during the idle period in burst-mode parallel optical links. The proposed architecture is similar to a conventional architecture presently implemented in data-centers. In the proposed technique, when the burst-mode channel is idle, it’s PR control code is updated by the always-on channel. Simulation results demonstrate a fast CDR lock time of 26 UIs, when it goes from an idle-state to its on-state, even in the presence of a ppm frequency offset between the clock and the data.

This work has been compared to the most relevant burst-mode works as presented in Table II. The proposed CDR, together with the presented sense circuit, takes less than 49 UIs to lock. However, the CDR takes only 26 UIs from the time it is turned on by the ON-OFF signal. In contrast, the clock recovery time presented in [5] and [6] is 352 UIs and 463 UIs, respectively from the time the CDR is powered on. The work of [8] concentrates on both timing and output DC-level recovery simultaneously and takes 58 UIs for the lock. This work achieves a faster lock time and a comparable power dissipation to [11] implemented in the same technology. The proposed work incorporates ILOs. Their lock time is the most significant factor in determining the overall turn-on time of the CDR. Reducing the lock time of the ILO will correspondingly reduce the lock time of the CDR.

The off-state power does not include the power of the always-on channel and the REF_CK, both of which can be amortized over the lanes. In a scenario with one always-on lane and 11 burst-mode lanes, the worst case off-state power (assuming no useful data over the always-on channel) of each burst-mode CDR becomes 2.35 mW as compared to 0.58 mW for a burst-mode CDR alone. This is still significantly smaller than the 19.5 mW dissipated by an always-on CDR.

Finally, the power overhead of the proposed design (2.6 %) is better compared to [8] (22 %) working at the same data rate. Therefore, the presented work, which has an area overhead of 597 µm² (1.3 % of the non-burst-mode area) coming from the “Off Detect” and the “Burst Detect” circuits, and an on-state power overhead of 0.5 mW, is an attractive solution for energy-efficient parallel optical links.

ACKNOWLEDGMENT

The authors thank CMC Microsystems for CAD tools, and Ted Obuchowicz for CAD tool support.

REFERENCES

[1] J. A. Tatum, L. A. G. D. Gazula, J. K. Guenter, R. H. Johnson, J. King, C. Kocot, G. D. Landry, I. Lyubomirsky, A. N. Maclnnes, E. M. Shaw, K. Balemnarth, R. Shubochkin, D. Vaidya, M. Yan and F. Tang, “VCSEL-Based Interconnects for Current and Future Data Centers,” J. Lightwave Technology, vol. 33, no. 4, pp. 727-732, Feb. 2015.
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2022.3162927, IEEE Access

[2] D. Abts, M. R. Marty, P. M. Wells, P. Klausler, and H. Liu, "Energy proportional datacenter networks," ACM 37th International Symposium on Computer Architecture (ISCA), Jun. 2010, pp. 338-347.

[3] A. Roy, H. Zeng, J. Bagga, G. Porter and A. C. Snoeren, "Inside the social network’s (Datacenter) network," ACM SIGCOMM Computer Communication Review, pp. 123-137, Aug. 2015.

[4] M. Dayarathna, Y. Wen and R. Fan, "Data Center Energy Consumption Modeling: A Survey," IEEE Communications Surveys & Tutorials, vol. 18, no. 1, pp. 732-794, Firstquarter 2016

[5] I. Ozkaya, A. Cevrero, P. A. Francesc, C. Menolfi, M. Braendli, T. Morf, D. Kuchta, L. Kull, M. Kossel, D. Luu, M. Meghelli, Y. Leblebici, and T. Toiil, “A 56Gb/s burst-mode NRZ optical receiver with 6.8ns power-on and CDR-Lock time for adaptive optical links in 14nm FinFET CMOS,” IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers, Feb. 2018, pp. 266–268.

[6] A. Rylyakov, J. Proesel, S. Rylov, B. Lee, J. Bulzacchelli, A. Ardey, B. Parker, M. Beakes, C. Baks, C. Schow, and M. Meghelli, “A 25Gb/s Burst-Mode Receiver for Rapidly Reconfigurable Optical Networks,” IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers, Feb. 2015, pp. 400–402.

[7] T. Morf, M. Seifried, A. Cevrero, I. Ozkaya, C. Menolfi, D. Kuchta, M. Kossel, P. Francesc, L. Kull, J. Kropp, and T. Toiil, “VCSEL-based optical links in burst-mode slow optical power ramp-up and how to achieve ultra-short wake-up times,” Electronics Letters, vol. 53, no. 19, pp. 1325-1327, Sep. 2017.

[8] A. K. M. D. Hossain, Aurangozab, and M. Hossain, “Burst mode optical receiver with 10 ns lock time based on concurrent DC offset and timing recovery technique,” IEEE J. Optical Communications and Networking, vol. 10, no. 2, pp. 65-78, Feb. 2018.

[9] T. Anand, M. Talegaonkar, A. Elkholy, S. Saxena, A. Elshaely, and P. K. Hanumolu, “A 7 Gb/s embedded clock transceiver for energy proportional links,” IEEE J. of Solid-State Circuits, vol. 50, no. 12, pp. 3101-3119, Dec. 2015.

[10] A. B. Mazwar, T. Kuriyama, and H. Ueda, “10 Gbps optical burst mode receiver with fast response and high stability,” IEEE International Conference on Communications (ICC), May 2016, pp. 1-6.

[11] D. Kim, M. G. Ahmed, W.-S. Choi, A. Elkholy and P. K. Hanumolu, “A 12-Gb/s 10 ns Turn-On Time Rapid ON/OFF Band-Rate DFE Receiver in 65-nm CMOS,” IEEE J. Solid-State Circuits, vol. 55, no. 8, pp. 2196–2205, Aug. 2020.

[12] A. I. Abbas, “Power-proportional optical links,” Ph.D. dissertation, ECE Dept., Concordia Univ., Montreal, QC, 2021.

[13] R. Kreienkamp, U. Langmann, C. Zimmermann, T. Aoyama and H. Siedhoff, "A 10-Gb/s CMOS clock and data recovery circuit with an analog phase interpolator," IEEE J. Solid-State Circuits, vol. 40, no. 3, pp. 736-743, Mar. 2005.

[14] A. Yuen, K. Giboney, E. Wong, L. Buckman, D. Haritos, P. Rosenberg, J. Stamicky, and D. Dolfi, “Parallel optical links for gigabyte/s data communication,” Symposium. IEEE Gallium Arsenide Integrated Circuit (GaAs IC) Symposium. Annual Technical Digest, Oct. 1997, pp. 193–196.

[15] K. Giboney et al., “Next-generation parallel-optical data links,” IEEE Lasers and Electro-Optics Society (LEOS), Nov. 2001, pp. 859-860.

[16] A. Agarwal, A. Liu, P. K. Hanumolu, and G.-Y. Wei, “An 8x5 Gb/s parallel receiver with collaborative timing recovery,” IEEE J. Solid-State Circuits, vol. 44, no. 11, pp. 3120–3130, Nov. 2009.

[17] Reflex Photonics Inc. “120 Gbps Pluggable/Surface Mount 12 Channel Parallel Fiber Optic Engines,” 970-00242 REV07 datasheet, 2016.

[18] A. Cevrero, I. Ozkaya, P. A. Francesc, C. Menolfi, M. Braendli, T. Morf, D. Kuchta, M. Kossel, L. Kull, D. Luu, J. Proesel, Y. Leblebici, and T. Toiil, "A 60 Gb/s 1.9 pJ/bit NRZ optical-receiver with low latency digital CDR in 14nm CMOS FinFET,” Symposium on VLSI Circuits, Jun. 2017, pp. 320-321.

[19] H. Wang and A. Hajimiri, “A wideband CMOS linear digital phase rotator,” IEEE Custom Integrated Circuits Conference (CICC), Sep. 2017, pp. 671-674.

[20] P. K. Hanumolu, G.-Y. Wei, and U.-K. Moon, “A wide-tracking range clock and data recovery circuit,” IEEE J. Solid-State Circuits, vol. 43, no. 2, pp. 425–439, Feb. 2008.

[21] L. Rodoni, G. V. Buren, A. Huber, M. Schmatz and H. Jackel, “A 5.75 to 44 Gb/s Quarter Rate CDR With Data Rate Selection in 90 nm Bulk CMOS,” IEEE J. Solid-State Circuits, vol. 44, no. 7, pp. 1927-1941, Jul. 2009.

[22] A. Ragab Y. Liu, K. Hu, P. Chiang, and S. Palermo, “Receiver jitter tracking characteristics in high-speed source synchronous links,” J. Electrical and Computer Engineering, vol. 2011, pp. 1-15, Nov. 2011.