I. INTRODUCTION

Today’s access networks represent around 70% of the overall power consumption of telecommunications networks, while core networks represent about 15% of it. However as the traffic on Internet increases, it is forecast that the power consumption in telecommunications networks will almost double by 2017, with access and core networks accounting for around 38% and 42% of such power consumption, respectively [1]. For these reasons, it is of capital importance to reduce the power consumption in also core networks.

WDM-based core optical networks have helped reducing the power consumption by allowing large amounts of traffic to bypass core routers through optical switching. Several research efforts have been focused on developing energy-efficient devices and network management strategies. A promising proposed solution consists in re-routing traffic carried by scarcely-filled wavelengths during low-load periods, in order to set some opto-electronic transmitters [i.e., transponders (TSP)] into a low-power mode or, eventually, turning them off. Authors in [2] introduced a novel TSP architecture whose novelty lies in the capability of putting the TSP modules in a low-power consumption mode (idle mode) when it does not support traffic; at the same time, the TSP modules can be turned fully operative (ON mode) in a short time, whenever incoming connections are supported. Specifically, the proposed TSP features three power states, described as follows:

1) Powered-on (ON) state: the TSP is receiving/transmitting traffic.

2) IDLE state: the TSP is set into a low-power mode (i.e., its power consumption is significantly lower than the consumption in the ON state) since it is not transmitting/receiving traffic. In this state, the TSP requires much shorter time to be activated compared to when the TSP is in the OFF state, as the components which require high stability (the lasers at the emitter and the photodiodes at the receiver) are left powered-on to avoid long activation times (on the other hand, the components whose activation time is not critical can be powered-off).

3) Powered-off (OFF) state: all the components of the TSP are powered-off.

Fig. 1 depicts the inter-state transitions. Let us now consider the case of different classes of connections [3]. In this context, the set-up time requirements of the connections might become a critical parameter. For instance, in multi-service networks, high-priority (C_h) traffic, typically requires short set-up times (few ms.). As a consequence it cannot be supported by TSPs in the OFF power state due to the long boot-up time (≈60 s. [2]). However, it can be supported by TSPs in the IDLE state, for which the transition to the ON state is negligible. On the other hand, low-priority (C_l) traffic can employ OFF TSPs as it can wait for longer boot-up times. Note that, by reserving some of the available TSPs per node for C_h traffic and putting them on IDLE state, it is possible to satisfy both C_h and C_l connections set-up time requirements, and, at the same time reduce the power consumption of the network. For instance, authors in [3]...
reported daily power savings around 56% with respect to traditional WDM network scenarios. However, besides the power savings, we must carefully evaluate the impact of this power management strategy on the blocking of the connections, e.g., due to the lack of available TSPs, specifically, lack of IDLE TSPs for \( C_h \) traffic and lack of OFF TSPs for \( C_l \) traffic.

Only few papers have evaluated analytically the blocking of the connections on WDM networks where multiple classes of services are considered (e.g., [4]–[8]). Additionally, the generic issue of resource reservation for \( C_h \) traffic has been addressed simulatively in some publications (e.g., [9]), however, with no reference to the power consumption issue. To the best of our knowledge, no paper has addressed the analytical evaluation of the blocking levels of the connections in the case some resources are reserved for \( C_h \) traffic and a TSP power management is used.

Moreover, a very important parameter to be evaluated is the wake-up time (transition from OFF to IDLE) of the TSPs, as it affects dramatically the overall blocking levels in a network, especially for \( C_h \) traffic. Therefore, the proper dimensioning of the number of TSPs to be put in the IDLE state is fundamental in order to achieve a good trade-off between blocking and power consumption. To analyze all these aspects, we present a novel Markov chain-based analysis in order to quantify the blocking levels of the aforementioned power management strategy and the associated power savings with respect to a traditional WDM network scenario. Moreover, the presented analytical model provides guidelines on how to choose the optimal value of the reserved TSPs for \( C_h \) traffic depending on multiple parameters such as the load, the wake-up time of the TSPs and the number of wavelengths per link. A previous, partial version of the presented analysis was presented in [10].

The rest of the paper is structured as follows: Section II describes the power management strategy considered here. Section III presents the analytical model to determine the blocking levels and the power savings. Section IV reports the numerical results obtained using the presented analytical model and discusses the trade-off between power savings and blocking levels. Finally, Section V draws the conclusion of the work.

II. POWER MANAGEMENT STRATEGY

In this section, we describe the power management strategy adopted and its main assumptions. The aim of the proposed strategy is to save the power consumed by unused devices by setting them into low-power mode (or, eventually, shutting them off), while still guaranteeing the set-up time target for the different classes (\( C_h \) and \( C_l \)) of connections. To avoid that \( C_h \) connections miss their set-up time target, we assume that a certain number (say \( k \)) of TSPs in a node, among those unused, are set in IDLE state and reserved for \( C_h \) connections. Assuming that every node in the optical network is equipped with \( N \) TSPs, with \( t \) representing the number of TSPs that are currently in the ON state, the other unused TSPs (i.e., \( N - k - t \)) are in OFF state and will be used for future incoming \( C_l \) connections. In our following model, we consider that each connection occupies an entire wavelength.\(^1\)

\(^1\)We consider a single line rate network, where TSPs operate at a single bit-rate (e.g., 100 Gb/s). In this context, a connection results from the aggregation at

Defining \( U \) as the current number of unused TSPs, we can say that, if \( U \geq k \), then exactly \( k \) are set in IDLE state. Otherwise, all the \( U \) TSPs are set into IDLE state. According to Fig. 1, when a \( C_h \) connection arrives, an IDLE TSP is activated to support the new \( C_h \) connection and an OFF TSP is woken-up (OFF to IDLE) such that the pool of \( k \) TSPs reserved for \( C_h \) traffic is maintained at the node. Such transition (from OFF to IDLE) can be performed if the number of OFF TSPs is greater than 0. By doing so, a \( C_h \) connection is blocked only if no IDLE TSPs are available. On the other hand, when a \( C_l \) connection arrives, an OFF TSP is booted-up. If no OFF TSP is available when a \( C_l \) connection arrives, then the connection is blocked, as the TSPs in IDLE state are reserved for the \( C_h \) connections. When a connection is torn-down, its TSPs can be either shut-down or hibernated (set into IDLE state) according to the number of available IDLE devices.

Note on the network architecture: In this work, an opaque network is considered, i.e., in every intermediate node an Optical-Electrical-Optical (O-E-O) conversion is performed (thus two WDM TSPs are employed for any link crossed by a lightpath). So every node is equipped with a pool of \( W \) TSPs on each incoming/outgoing link, which is considered as bi-directional. We focus on this scenario: 1) as the bit-rate of demands and the use of more complex modulation formats increases, the reach of the connections is no longer compatible with optical paths but with links, hence transparent networks will be more and more difficult to realize [11]; 2) since client demands have smaller sizes with respect to such channel capacities, for better filling the tunnels electrical grooming will be required; 3) thanks to the O-E-O conversion, the wavelength continuity constraint can be relaxed and serve more connections with the available wavelengths’ set (this situation tends to be preferable when the spectrum is scarce). For all these reasons, we believe that the derivation of an analytical model for the opaque scenario is still worth and can rise the interest of the readers. Nonetheless, in [12] we evaluated the performance of the presented power management strategy in translucent networks, to which the interested reader can refer.

III. BLOCKING ANALYSIS

In this section, we provide an analytical model for the evaluation of the blocking and power savings when adopting the power management strategy previously described. To facilitate the comprehension of the analysis, we start with a single network node case. Then, we introduce in the analysis the wake-up time parameter to derive its impact on the blocking and power savings. Finally, we extend the analysis to a whole network scenario.

A. Single Node Model (Without Wake-Up Time)

In this first step, we consider that the transition between the OFF and the IDLE state is instantaneous (wake-up time = 0). The main assumptions and variables that we consider are as stated below:

the source node from various client connections coming from an access/metro network which require the communication through a core network with a geographically distant node in the same destination access/metro network. Hence, the entire capacity of a wavelength is filled.
1) Connections arrive at the node following a Poisson process. The mean arrival rates for $C_h$ and $C_l$ connections are denoted as $\lambda_h$ and $\lambda_l$ respectively.

2) The duration of all the connections is assumed to be exponentially distributed with mean $\mu$.

3) The node is equipped with $N = (W \times$ nodal degree) WDM TSPs.

4) Among the TSPs in the node, $k$ (maximum) are reserved for $C_h$ connections and are put in IDLE state. We remind the reader that, if at least $k$ TSPs are actually unused, exactly $k$ TSPs are put in IDLE state. Otherwise, all unused TSPs are put in IDLE state and reserved for $C_h$ connections.

5) $\sigma_{x,y}$ represents the state in which the node has exactly $x$ TSPs in OFF and $y$ TSPs in IDLE. Note that the number of TSPs that are actually in ON can be calculated as $N - x - y$. Additionally, we define the probability of being in this state as $\pi_{x,y}$.

With such a notation, the evolution of the power-state of the TSPs within the node can be represented as a time-homogeneous finite-space Markov chain. Fig. 2 depicts the state diagram of the mentioned Markov chain, along with the blocking states for both classes of connections. Note that $C_h$ connections only experience blocking when all the TSPs in the node are in use, while $C_l$ connections experience blocking when no OFF TSPs are available (only IDLE TSPs are available).

Such Markov chain describes a birth–death process, for which we can obtain the state probabilities. For the sake of simplicity, we will denote with $s_i$, $0 \leq i \leq N$, the state where $N - i$ TSPs are in ON state, that is, $i$ TSPs are currently unused, and with $p_i$ the probability of such state. More specifically, $s_i$ and $p_i$ can be expressed as

$$s_i = \begin{cases} 
\sigma_{0,i}, & \text{if } 0 \leq i < k \\
\sigma_{i-k,k}, & \text{if } k < i \leq N 
\end{cases}$$

$$p_i = \begin{cases} 
\pi_{0,i}, & \text{if } 0 \leq i < k \\
\pi_{i-k,k}, & \text{if } k < i \leq N. 
\end{cases}$$

Then the state probabilities can be evaluated as

$$p_{N-1} = \frac{(\lambda_h + \lambda_l)}{\mu} \cdot p_N$$

$$p_{N-2} = \frac{2\lambda_l}{2\mu} \cdot p_{N-1} = \frac{(\lambda_h + \lambda_l)^2}{2\mu^2} \cdot p_N$$

$$p_k = \frac{(\lambda_h + \lambda_l)^{N-k}}{(N-k)!\mu^{N-k}} \cdot p_N$$

and can be rewritten in closed form as

$$p_i = \begin{cases} 
\frac{(\lambda_h + \lambda_l)^{N-i}}{(N-i)!\mu^{N-i}} \cdot p_N, & \text{if } k \leq i \leq N \\
\frac{(\lambda_h + \lambda_l)^{N-k} \lambda_h^{k-i}}{(N-k)!\mu^{N-i}} \cdot p_N, & \text{if } 0 \leq i \leq k-1. 
\end{cases}$$

It can be seen that the state probabilities depend on $p_N$, which can be computed as

$$p_N = 1 - \sum_{i=0}^{N-1} p_i.$$  \hspace{1cm} (3)

Given the state probabilities computed using Eqs. (2) and (3), we can evaluate the blocking probabilities of both $C_h$ and $C_l$ traffic and the total blocking probability, (respectively $P_b,h$, $P_b,l$ and $P_{b,tot}$) as

$$P_{b,h} = p_0, \quad P_{b,l} = \sum_{i=0}^{k} p_i, \quad P_{b,tot} = h \cdot P_{b,h} + l \cdot P_{b,l}$$ \hspace{1cm} (4)

where we denote as $h$ and $l$ the share of $C_h$ and $C_l$ connections, respectively (0 ≤ $h$, $l$ ≤ 1; $h + l = 1$).

Moreover, the average power consumption in the node can be also estimated. Given the state probabilities, we can determine the average number of TSPs that are in IDLE and ON state, respectively, as the following, where $a \land b$ represents the minimum between two values:

$$\text{TSP}_{\text{idle}} = \sum_{i=0}^{N} (i \land k) \cdot p_i, \quad \text{TSP}_{\text{on}} = \sum_{i=0}^{N} (N - i) \cdot p_i.$$ \hspace{1cm} (5)

Using (5), the average power consumption of the node can be written as

$$\bar{P}_{\text{power}} = P_{\text{idle}} \cdot \text{TSP}_{\text{idle}} + P_{\text{on}} \cdot \text{TSP}_{\text{on}}$$ \hspace{1cm} (6)

where $P_{\text{idle}}$ and $P_{\text{on}}$ are the power consumption of a TSP being on the IDLE and the ON state respectively.

**B. Single Node Model (With Wake-Up Time)**

The previous model assumed that the transition between the OFF and the IDLE state is instantaneous. Actually, in a more realistic scenario, this transition time may not be negligible and its duration may impact the performance of the power-management strategy.

Indeed, every time an IDLE TSP is activated due to the arrival of a $C_h$ connection, if there are spare OFF TSPs, an OFF TSP will be woken-up and put on IDLE. However, during a wake-up time (from now on referred as $T_{\text{W}}$), and depending on the value of $k$, it can happen that, at certain point, new $C_h$ connections are blocked due to the lack of IDLE TSPs. To capture the influence of $T_{\text{W}}$ on the blocking probability and the average power...
consumption, we generalize the Markov chain presented in the previous section. To this end, in addition to the variables and definitions stated in the previous section:

1) We define $\omega_t$ as the frequency of waking-up a TSP, that is, $\omega_t = \frac{1}{T_w}$. To the best of our knowledge, there is no work providing an analysis of the statistical distribution of the wake-up time of a sleep-enabled TSP. In such circumstances, we assume the wake-up time to be exponentially distributed, since, in lack of further knowledge, it is the most conservative statistical distribution consistent with its mean value[14].

2) Since we are targeting a circuit-switched network, where connections typically have very long holding times, we assume that $T_w$ is much shorter than the average duration of the connections, that is, $T_w \ll \frac{1}{\mu}$. Conversely, $\mu \ll \omega_t$. Hence, in a realistic network scenario, the probability of a connection terminating during the duration of a wake-up operation is negligible and such transition is not considered in the current analysis. Nevertheless, we will assess the impact of this assumption on the estimation of the blocking probability later on during Section IV.

Also in this case, the evolution of the power-state of the TSPs within the node can be described as a time-homogeneous finite-space Markov chain, with the difference that the introduction of the wake-up time variable leads to a bi-dimensional Markov chain. Additionally, in this case, we follow a slightly different notation for the states and their probabilities. Due to the presence of TSPs that are operating a power transition from OFF to IDLE (wake-up operation), the state $\sigma_{x,y}$ now represents the state where the total number of OFF TSPs plus the ones waking-up is equal to $x$, while the total number of IDLE TSPs is equal to $y$. That is, the right number represents the number of TSPs that are on IDLE, while the left number represents the summation of the TSPs that are in OFF state and the ones that are waking-up, i.e., performing a transition from OFF to IDLE. More specifically, the actual number of TSPs that are waking-up in every state is equal to $x \land (k - y)$ while the actual number of TSPs that remain in the OFF state is $x - [x \land (k - y)]$. For example, the first row of the first column represents the state where $N - k$ TSPs are on OFF, 0 are waking-up and $k$ are on IDLE, while the last row of the first column represents the state where $N - 2k$ TSPs are on OFF, $k$ are waking-up and 0 are on IDLE. Also, note that in some states the maximum number of TSPs that are waking-up is not $k$ but a number equal to all the OFF TSPs before a wake-up operation is triggered. For instance, all states from the third to last row of the sixth column, represent the situation where zero OFF and two waking-up TSPs are present, with the only difference between them being the number of IDLE TSPs (the value of $y$). Note that the use of two indexes is sufficient to represent all considered power states in the Markov chain, allowing us to formulate a more compact and concise analytical model than the one that would result if three indexes (OFF, waking-up and IDLE) were used.

Following this notation for the states, Fig. 3 depicts the state diagram of the Markov chain, along with the blocking states for both classes of connections. The first row of the Markov chain is the same as in the previous case, while the extra rows represent the states where an IDLE TSP has been used and an OFF TSP is waking-up. As before, $C_h$ connections are blocked when all the TSPs on the node are used and $C_l$ connections are blocked when no OFF TSPs are available. However, now new blocking states appear for both classes due to the influence of $T_w$.

In the new blocking states for $C_h$ (the ones in the lowest row of the chain), connections are blocked due to the exhaustion of IDLE TSPs due to multiple arrivals of $C_h$ connections during a
After a $C_h$ connection arrives, whenever possible, a wake-up operation is performed in order to maintain a pool of $k$ IDLE TSPs. If during this operation, $k - 1$ $C_h$ connections arrive to the node, all the IDLE TSPs will become exhausted, so the next $C_h$ connection will be blocked (see Fig. 4).

As for $C_l$ connections, the new blocking states are in the right diagonal part of the Markov chain. In those states, all the current OFF TSPs are operating a power transition from OFF to IDLE, hence, they are unavailable for $C_l$ connections. Specifically, all OFF TSPs are operating a power transition in the states where $1 \leq x \leq k - y; 0 \leq y \leq k - 1$. When this happens, new incoming $C_l$ connections are blocked since there are no OFF TSPs that can be used to set-up them. Note that in such states, the pool of IDLE TSPs after all wake-up operations have ended will not be equal to $k$ but instead will be equal to $x + y$.

We can now calculate the state probabilities. For this, we represent as $p_{m,n}$ the probability of state $\sigma_{x,y}$ (for the sake of conciseness, we will only depict the final closed form of the probabilities), as shown (7), at the bottom of this page, with

\[
\lambda_{m,n}^* = \begin{cases} 
\lambda_h + \lambda_l, & \text{if } k - n + 1 \leq m \leq N - k; 1 \leq n \leq k \\
\lambda_h, & \text{if } 0 \leq m \leq k - n; 1 \leq n \leq k \\
\lambda_l, & \text{if } k + 1 \leq m \leq N - k; n = 0 \\
0, & \text{if } 0 \leq m \leq k; n = 0 
\end{cases}
\]

\[
\lambda_{l,m}^* = \begin{cases} 
\lambda_l, & \text{if } k - n + 1 \leq m \leq N - k; 0 \leq n \leq k \\
0, & \text{if } 0 \leq m \leq k - n; 0 \leq n \leq k. 
\end{cases}
\]

Essentially, the state probabilities in the first row depend on the previous state at their left and on all states diagonally reaching this state. As for the states in the rest of the rows, their probabilities depend on all neighboring states that have an incident transition to that state. Additionally, as before, $p_{N-k,k}$ can be calculated as

\[
p_{N-k,k} = 1 - \sum_{m=0}^{N-k} \sum_{n=0}^{N-k-1} p_{m,n} - \sum_{m=0}^{N-k-1} p_{m,k}. \tag{10}
\]

With the values of the state probabilities computed using equations (7)–(10), we can evaluate the blocking probabilities of both $C_h$ and $C_l$ and the total blocking probability as

\[
P_{b,h} = \sum_{m=0}^{N-k} p_{m,0}, \quad P_{b,l} = \sum_{m=0}^{k-m} \sum_{n=0}^{N-k} p_{m,n}, \quad P_{b,tot} = h \cdot P_{b,h} + l \cdot P_{b,l}. \tag{11}
\]

We can also determine the average number of TSPs that are in IDLE ($\text{TSP}_{\text{idle}}$), ON ($\text{TSP}_{\text{on}}$) and the ones that are operating a power transition from OFF to IDLE ($\text{TSP}_{\text{trans}}$)

\[
\text{TSP}_{\text{idle}} = \sum_{m=0}^{N-k} \sum_{n=0}^{m} \pi_{m,n} \cdot p_{m,n},
\]

\[
\text{TSP}_{\text{on}} = \sum_{m=0}^{N-k} \sum_{n=0}^{m} (N - m - n) \cdot p_{m,n},
\]

\[
\text{TSP}_{\text{trans}} = \sum_{m=1}^{N-k} \sum_{n=0}^{m} (m \land (k - n)) \cdot p_{m,n}. \tag{12}
\]

Using (12), we can calculate the average power consumption of the node, taking into account that current devices are capacitance-based and their power consumption can be considered lineal. Hence, we can suppose that the average power during transitions can be calculated as half of the difference between the two states

\[
P_{\text{power}} = P_{\text{idle}} \cdot \text{TSP}_{\text{idle}} + P_{\text{on}} \cdot \text{TSP}_{\text{on}} + \frac{1}{2} P_{\text{idle}} \cdot \text{TSP}_{\text{trans}}. \tag{13}
\]
C. Network-Wide Model

A more realistic scenario entails the analysis not only on a single node but on the whole network. In this case, the availability of the TSPs must be evaluated along the entire end-to-end path between source/destination nodes, where multiple connections interfere as they compete for the same resources.

To evaluate the blocking probability in a network, the well-known method called reduced load Erlang fixed point approximation method is used [13]. Due to the complexity of determining such analytical expressions, we consider that only the shortest-path route can be exploited between a source and a destination node. We consider this assumption acceptable as our main purpose is to provide an analytical model which is sufficiently accurate to evaluate the benefits of the TSP power management strategy and help to set the most effective number of IDLE TSPs per node when such strategy is employed (i.e., the value of $k$). The analysis of more complex scenarios is left for future work.

A first difference when compared to the single node case is that in the network model, the $N$ parameter is no longer used. Instead, we refer to the total number of wavelengths per fiber link, denoted as $W$. As we are considering an opaque network, every wavelength channel on a link entails the use of two TSPs, one at each end of a link. Moreover, we denote as $k$ the number of wavelengths that are being reserved for $C_h$ connections, thus, equivalently, at every end of a link, $k$ TSPs are being reserved for such connections. This means that, for a $C_h$ connection to be routed, in every link that composes the end-to-end route, at least one wavelength has to be available, while for $C_l$ connections to be routed, at least $k + 1$ wavelengths have to be available.

Let us discuss in detail the analytical model for such scenario. It is based on the analytical approach presented in [5], for which we utilize the independence variation as a baseline of our work, since in an opaque network the load correlation introduced by the wavelength continuity constraint does not apply. We first introduce the parameters and the corresponding definitions that will be used in our analysis:

1) $G = (V, E)$ is the physical topology of the network, consisting of a set $V$ of nodes and a set $E$ of bi-directional links.
2) $R$ is the set of connection requests between source-destination pairs.
3) We define $r_{sd}$ as the route from source node $s$ to destination node $d$.
4) $H_{ij}$ and $L_{ij}$ are the random variables which represent the number of wavelengths on link $(i, j) \in E$ that can be used to set-up $C_h$ and $C_l$ connections, respectively.
5) $q_{m,n}^{ij}$ is the probability of having exactly a total of $m$ wavelengths that their associated TSPs are OFF (hence, usable by $C_l$ connections) or performing a wake-up operation and a total of $n$ wavelengths that their associated TSPs are ON (hence, usable by $C_h$ connections). Note that in the network-wide model we follow the same notation to represent the state of a link as the one introduced during Section III-B to represent the state of a node. Note also that, with such notation, $H_{ij} = n$ while $L_{ij} = m - [m \land (k - n)]$.
6) $X_{sd}$ and $Z_{sd}$ are the random variables which represent the number of wavelengths on the whole route between a source/destination pair $(s, d) \in R$ that can be used to set-up $C_h$ and $C_l$ connections, respectively.
7) $\lambda_{h, sd}$ and $\lambda_{l, sd}$ are the base arrival rates of $C_h$ and $C_l$ connections between nodes $s$ and $d$.
8) $\alpha_{m,n}^{ij}$ is the connection arrival rate at link $(i, j)$ given that the combination of wavelengths that are in the OFF state plus the ones that are in the “wake-up” state is exactly equal to $m$ and $n$ wavelengths are in the IDLE state, with $\alpha_{m,n}^{h,ij}$ and $\alpha_{m,n}^{l,ij}$ the contributions for each kind of traffic. These variables are the so called state-dependent arrival rates [15] and capture the influence of every connection on the others.
9) $P_{b, sd}$ is the blocking probability for connections from $s$ to $d$.

With such definitions, the availability of wavelengths over a particular link $(i, j)$ can be represented employing the same Markov chain depicted in Fig. 3, as the wavelengths in the links follow the same state transitions since their particular availability is strictly related to the availability of TSPs at nodes. Following this consideration, the equations used to compute the state probabilities in Section III-B can be used now, with the proper variable substitutions: the probabilities $p_{n,m}$ are now substituted by $q_{m,n}^{ij}$, parameter $N$ is now substituted by $W$ and the arrival rates are no longer constant but they depend on the state of the link. This last aspect is the main difference in the network-wide model and is what we defined as state-dependent arrivals, represented with variables $\alpha_{m,n}^{ij}$. Therefore, taking into account this consideration, and using the same analysis presented in Section III-B, we can determine the state probabilities (the $q$ variables) for each one of the links $(i, j)$ as shown (14), at the bottom of the next page.

Similarly to the single node case, $q_{W-k,k}^{ij}$ can be calculated as

$$q_{W-k,k}^{ij} = 1 - \sum_{m=0}^{W-k-1} \sum_{n=0}^{W-k} q_{m,n}^{ij} - \sum_{m=0}^{W-k} q_{m,k}^{ij}.$$  \hspace{1cm} (15)

Once we know the state probabilities for every link, we can evaluate the blocking for every connection belonging to $R$ as

$$P_{b, sd} = h \cdot (1 - Pr\{X_{sd} > 0\}) + l \cdot (1 - Pr\{Z_{sd} > 0\}).$$ \hspace{1cm} (16)

The terms $Pr\{X_{sd} > 0\}$ and $Pr\{Z_{sd} > 0\}$, i.e., the probabilities of having enough resources along the end-to-end route to serve a $C_h$ and a $C_l$ connection, respectively, can be computed as the product of the probabilities of having enough resources in all links along the route, since it is necessary that all links from source to destination have enough free resources to serve the connections

$$Pr\{X_{sd} > 0\} = \prod_{(i,j) \in r_{sd}} Pr\{H_{ij} > 0\}$$

$$Pr\{Z_{sd} > 0\} = \prod_{(i,j) \in r_{sd}} Pr\{L_{ij} > 0\}$$ \hspace{1cm} (17)

where $Pr\{H_{ij} > 0\}$ and $Pr\{L_{ij} > 0\}$ are the probabilities of having enough resources in the link $(i, j)$ to serve a $C_h$ and a $C_l$.
connection, and are obtained as follows considering the blocking states depicted in Fig. 3

\[
Pr\{H_{ij} > 0\} = 1 - \sum_{m=0}^{W-k} q_{m,0}^i \cdot Pr\{L_{ij} > 0\}
\]

\[
= 1 - \sum_{m=0}^{k} \sum_{n=0}^{k-m} q_{m,n}^{ij}.
\]

(18)

Hence, the blocking probability in the network can be evaluated as

\[
P_{b,tot} = \sum_{(s,d) \in R} t_{sd} P_{b, sd}
\]

(19)

where \( t_{sd} \) represents the share of traffic from \( s \) to \( d \) respect the total traffic in the network.

We can estimate the average power consumption in the network following the same method as in the single node case

\[
TSP_{ij, idle} = 2 \cdot \sum_{m=0}^{W-k} \sum_{n=0}^{k} n \cdot q_{m,n}^{ij}
\]

\[
TSP_{ij, on} = 2 \cdot \sum_{m=0}^{W-k-1} \sum_{n=0}^{k-1} (W - m - n) \cdot q_{m,n}^{ij}
\]

\[
TSP_{ij, trans} = 2 \cdot \sum_{m=0}^{W-k-1} \sum_{n=0}^{k-1} (m \land (k - n)) \cdot q_{m,n}^{ij}
\]

\[
P_{ower} = \sum_{(i,j) \in E} \left( P_{idle} \cdot TSP_{ij, idle} + P_{on} \cdot TSP_{ij, on} + \frac{1}{2} P_{idle} \cdot TSP_{ij, trans} \right)
\]

(20)

where the factor 2 is due to the fact that each link entails the use of a TSP on each side of the link.

However, in the previous analysis, the state probabilities depend on the state-dependent arrival rates, which have to be still evaluated. The process to evaluate them is an iterative process, which involves knowing the blocking on each link, and the steps are as follows. Denoting as \( \alpha_{sd,i} \) the arrival rate of connections \( (s,d) \in R \) at link \( (i,j) \) given that the combination of wavelengths that are in the OFF state plus the ones that are in the “wake-up” state is exactly equal to \( m \) and \( n \) wavelengths are in the IDLE state, its value is as follows (21), at the bottom of the page.

The probability terms in (21) are computed taking into account the availability of resources on all links along the end-to-end route. Hence, it is necessary to consider the probability of the states that have at least some resources to allow the establishment of \( C_h, C_l \) or both types of connections. The resulting equations are

\[
Pr\{X_{sd} > 0\} = m - [m \land (k - n)]; H_{ij} = n = \prod_{(r,s) \in E(i,j)} \left( 1 - \sum_{u=0}^{W-k} q_{ru,0}^{ry} \right),
\]

if \( 0 \leq m \leq k - n; 1 \leq n \leq k \)

(22)

\[
Pr\{Z_{sd} > 0\} = m - [m \land (k - n)]; H_{ij} = n = \prod_{(r,s) \in E(i,j)} \left( 1 - \sum_{u=0}^{k-n} q_{ru,v}^{rv} \right),
\]

if \( k + 1 \leq m \leq W - k; n = 0 \)

(23)

\[
\alpha_{sd,ij} = \begin{cases} 
0, & \text{if } 0 \leq m \leq k; n = 0 \\
\lambda_{h, sd} \cdot Pr\{X_{sd} > 0\} = m - [m \land (k - n)]; H_{ij} = n, & \text{if } 0 \leq m \leq k - n; 1 \leq n \leq k \\
\lambda_1 \cdot Pr\{Z_{sd} > 0\} = m - [m \land (k - n)]; H_{ij} = n, & \text{if } k + 1 \leq m \leq W - k; n = 0 \\
(\lambda_h \cdot \lambda_l) \cdot Pr\{X_{sd} > 0\}; Z_{sd} > 0\} = m - [m \land (k - n)]; H_{ij} = n, & \text{if } k - n + 1 \leq m \leq W - k; 1 \leq n \leq k.
\end{cases}
\]

(21)
\[
\Pr\{X_{sd} > 0; Z_{sd} > 0|L_{ij} = m - [m \wedge (k - n)]; H_{ij} = n\} = \prod_{(x,y)\in R_{sd}, (x,y)\neq (i,j)} \left(1 - \sum_{u=0}^{W-k} q_{u,0} - \sum_{u=0}^{k} \sum_{v=0}^{k-n} q_{u,v}\right),
\]

if \(k - n + 1 \leq m \leq W - k; 1 \leq n \leq k\).

Therefore, using equations (21)-(24), the new value of the state-dependent arrival rates can be computed as the summation of all traffic that goes through link \((i, j)\)

\[
\alpha_{ij}^{sd} = \sum_{(s,d): (i,j) \in r_{sd}} \alpha_{m,n}^{sd,ij}.
\]

Once all involved variables have been correctly evaluated, we can determine the blocking levels in the network through an iterative process. First, by setting the state-dependent arrivals to an initial value, the blocking on each link can be obtained. When determined, it will be used to calculate the value of the state-dependent arrivals for the next iteration, which, in turn, will allow us to evaluate the blocking in each link for the next iteration and so on. Through this process, we can iteratively calculate the blocking probability for all pairs \((s, d) \in R\) until a desired maximum error \(\epsilon\) between two consecutive iterations is achieved. This maximum error is used to decide when the iterative process has converged and the obtained blocking probability has reached a steady value, so further iterations would not change significantly the obtained value. Note that if a high value of \(\epsilon\) is chosen, it may happen that the iterative process terminates before a sufficient number of iterations have been executed, so the obtained value could be far from the real blocking probability. Hence, to obtain accurate results, a sufficiently small value of \(\epsilon\) has to be chosen. The steps of the calculation are the followings:

1) Set the value of the maximum allowed error between two consecutive iterations (e.g., \(\epsilon = 10^{-6}\)). Initialize the estimated blocking probability, for every pair \((s, d) \in R\), to zero, i.e., \(\hat{P}_{b,sd} = 0, \forall (s,d) \in R\). For all links \((i, j) \in E\) initialize the state-dependent arrival rates as

\[
\alpha_{m,n}^{sd,ij} = \begin{cases} 
0, & \text{if } 0 \leq m \leq k; n = 0 \\
\sum_{(s,d): (i,j) \in r_{sd}} \lambda_{h,sd}, & \text{if } 0 \leq m \leq k - n; 1 \leq n \leq k \\
\sum_{(s,d): (i,j) \in r_{sd}} \lambda_{l,sd}, & \text{if } k + 1 \leq m \leq W - k; n = 0 \\
\sum_{(s,d): (i,j) \in r_{sd}} (\lambda_{h,sd} + \lambda_{l,sd}), & \text{if } k - n + 1 \leq m \leq W - k; 1 \leq n \leq k.
\end{cases}
\]

2) Compute the state probabilities for every link using Eqs. (14)-(15).

3) Calculate the state-dependent arrival rates for all links according to Eqs. (21)-(25).

4) Compute, for every couple \((s, d) \in R\), the blocking probability \(P_{b,sd}\) using Eqs. (16)-(18). If \(\max_{(s,d)} |\hat{P}_{b,sd} - \hat{P}_{b,sd}| < \epsilon\) then terminate and evaluate the total blocking and average power consumption according to (19) and (20), respectively. Otherwise, set \(\hat{P}_{b,sd} = P_{b,sd}\) and go back to step 2.

IV. CASE STUDIES AND RESULTS

A. Scenario Description

In this section, we evaluate the proposed power management strategy through a set of numerical results. All the results in this section have been obtained using the COST239 network topology (11 nodes and 26 bidirectional links) [16]. We assume that all the traffic is uniformly distributed among all pairs \((s, d)\).

Additionally, we assume that the average duration of the connections is equal to 1, i.e., \(\mu = 1\), hence the load is equal to the average arrival rate of connections to the network \(\lambda\), with \(\lambda_{sd} = \lambda/R\) \(\forall (s,d) \in R\).

We have evaluated different scenarios, where the number of TSPs reserved for \(C_h\) connections \((k)\) is varied. Moreover, also the percentages of \(C_h\) and \(C_l\) connections over the total are varied. Note that varying these percentages has a direct effect on the values of the base arrival rates \(\lambda_h\) and \(\lambda_l\), i.e., given the overall connection arrival rate \(\lambda = \lambda_h + \lambda_l\), we can gather the values of \(\lambda_h\) and \(\lambda_l\) considering that \(h + l = 1\), \(h = \lambda_h/\lambda\) and \(l = \lambda_l/\lambda\). Additionally, we have evaluated multiple scenarios with different values of \(T_w\) (versus \(\omega_t\)). In all our tests, we have set \(\epsilon = 10^{-12}\). Moreover, we set \(P_{on} = 351\) W and \(P_{idle} = 18\) W for the TSPs as in [2].

B. Model Evaluation

Fig. 5 shows, for increasing values of the network load, the total blocking probability with \(h = l = 0.5\), \(W = 32\) and \(T_w = 1\) ms (\(\omega_t = 1000\)). Additionally, in order to check the accuracy of the analytical model, we have also performed some simulations using an event-driven simulator considering the same power management strategy and network scenario as in the analytical model. All the simulations have been performed considering \(4 \times 10^6\) connection arrivals to the network. All simulation results have been obtained with 95% confidence intervals lower than one order of magnitude respect the average value (the size of the confidence interval is around a relative 2.8% in the worst case and around a relative 1% in average), assuring the statistical accuracy of the presented results. In addition to various values of \(k\), we have tested two special cases as benchmarks, namely:

1) all TSPs are powered-on, that is, any TSP can be used by either \(C_h\) or \(C_l\) connections (“All-On” in the figures).

2) unused TSPs are powered-off, that is, no TSPs are set in IDLE state, hence, all the \(C_h\) connections are blocked, so the available capacity is completely exploited by \(C_l\) connections (“\(k = 0\)” in the figures).
In Fig. 5, we can see that for low values of $k$, almost no performance degradation is experienced in terms of blocking probability, while appreciable performance degradations arise for $k \geq 8$. However, if we observe the case $k = 1$, we can see a sudden degradation of the blocking probability for low loads. Such degradation is due to the influence of the wake-up time and will be commented in details later on in Section IV-C. Additionally, it can be appreciated that the differences between the analytical model and the simulations are very small (less than a relative 3% deviation) with almost identical blocking figures. Such differences arise from the fact that the analytical model does not take into account the possibility of connections terminating during a wake-up operation, while such transitions are present in the simulations. Note, however, that from the depicted results, the impact of this assumption is very small, demonstrating the correctness and accuracy of the analytical model. As for the blocking probability of the two different classes, the general trend is that the blocking probability of the $C_h$ traffic decreases with increasing values of $k$ while the blocking probability of $C_l$ traffic increases. This is due to the fact that higher values of $k$ imply that more TSPs are put on idle and reserved for $C_h$ traffic, so less resources can be exploited by $C_l$ traffic. As an example, taking a load equal to 300, for $k = 1$, the blocking of $C_h$ and $C_l$ traffic are 0.85% and 0.06%, respectively, while for $k = 2$ they are 0.01% and 0.12% and for $k = 4$ they are 0.00185% and 0.31%.

We also evaluated the average and the normalized power consumption for the analytical model. The normalized power consumption is computed as

\[
\text{normalized power} = \frac{\text{total power}}{\text{n. of connections} \cdot (1 - P_b)}.
\]  

(27)

In Fig. 6, it can be observed that power savings of at least 32% can be obtained by setting some resources in IDLE state with respect to the All-On scenario. Moreover, huge savings, around 80%, are observed for lower loads and for all the values of $k$. The impact of $k$ becomes relevant for higher loads, when reserving more resources for $C_h$ connections affects the power consumption values, mainly due to the fact that higher blocking is experienced for increasing $k$, as more $C_h$ are being blocked.
In order to remove the influence of blocking from the power consumption results, we introduced the normalized power metric. We can observe that, when the blocking is negligible, the power consumption increases with \( k \), as more unused resources are maintained in IDLE state unnecessarily. On the other hand, for increasing loads, the normalized power consumption decreases more rapidly for higher values of \( k \). This is due to the fact that, for higher values of \( k \), blocking mainly occurs for multiple-hops connections, as the chances of having enough available TSPs in every hop decrease with the length of the path. Thus, the connections accepted for higher \( k \) are typically single-hop and consume less power than those accepted for lower values of \( k \).

C. Influence of the Wake-Up Time

Although, as mentioned before, low values of \( k \) do not entail significant performance degradations in terms of blocking, in Fig. 5 we can observe a sudden degradation in the \( k = 1 \) (for loads below 300 Erlangs) performance. This is an effect of the wake-up time of the TSPs. To further investigate this effect, we focused on the case \( k = 1 \). Fig. 7 depicts the blocking probability as a function of the load for various values of \( T_w \). Moreover, we also plot the case where the \( T_w \) is not considered, as presented in [10], denoted as “No \( T_w \)”. As expected, the value of the \( T_w \) has a substantial impact on the blocking probability. Blocking increases for increasing values of \( T_w \). This is due to the effect mentioned in Section III: after an IDLE TSP is used, and if there are still spare OFF TSPs, a wake-up operation is performed in order to maintain the pool of \( k \) IDLE TSPs; if during this wake-up operation, \( k - 1 \) \( C_h \) connections arrive, all the IDLE TSPs will become exhausted, so the next \( C_h \) connections will be blocked. Intuitively, this situation is more likely to happen for longer \( T_w \). In fact Fig. 7 shows that higher values of \( T_w \) increase substantially the levels of blocking while for low values, its effect is almost negligible. Note also that for low loads, curves change their concavity. This can be explained as, for low loads, the main component of the blocking is the lack of IDLE TSPs during wake-up times, while for increasing loads, as connections start to compete with each other, the main component of the blocking is the lack of free resources. As for the case where the \( T_w \) is neglected, it can be seen that, indeed, not considering its influence can lead to substantial differences in the values of \( P_b \) when the load is small.

In light of the results depicted in Fig. 7, it is clear that the correct dimensioning of \( k \) to obtain a certain network performance (say, \( P_b \leq 5\% \)) depends highly on the value of \( T_w \) as well as on the share of \( C_h \) connections (since they are the connections that are affected by the wake-up time). To this end, we have conducted additional studies where, for various values of \( T_w \) and \( h \), we varied \( k \) in order to determine the most appropriate value of \( k \) to achieve a blocking not greater than 5% for a reference load of 500 Erlangs. The obtained results are depicted in Fig. 8, where the black solid line marks the aforementioned threshold for the \( P_b \).

It can be seen that for the selected load, even for low shares of \( C_h \) traffic, it is not possible to fulfill a 5% threshold on the \( P_b \) with a wake-up time of 100 ms, as it is still too long to avoid the lack of TSPs during a \( T_w \). In this case, it would be necessary to dimension the network for another working load (e.g., by increasing the number of wavelengths per link) or employ other technologies for the TSPs that allow lower wake-up times. We can see that, if the wake-up time is lower (e.g., 10 ms), as the share of \( C_h \) connections increases, it is necessary to reserve more IDLE TSPs, e.g., \( k = 1 \) for \( h = 0.2 \), \( k = 2 \) for \( h = 0.5 \) and \( k = 3 \) for \( h = 0.8 \). Moreover, it can be appreciated that, for increasing values of \( h \) the \( P_b \) curves have a minimum. This is due to the fact that, as the value of \( k \) increases, the blocking due to the lack of IDLE TSPs decreases but, at the same time, the blocking due to lack of resources for \( C_l \) connections increases, as less and less resources can be used by them.

V. Conclusion

In this paper, we provided a novel blocking model to evaluate a power management strategy for WDM TSPs based on setting unused TSPs into low-power idle state or powered-off in multi-service optical networks. Our model quantifies the trade-off between the power savings which can be obtained with this method and the performance degradation, when compared to the case with all TSPs always powered-on and without resources reservation. The results demonstrated at least 32%, and up to 80%, of power savings, for acceptable blocking probability values.

Moreover, we showed the impact over network blocking produced by the parameters considered in the TSP power.
management strategy, such as the number of TSPs reserved for high-priority connections, the wake-up time of the TSPs and the percentage of high-priority connections out of the total connections arrival rate. Therefore, the presented model can serve to provide guidelines on the dimensioning of an optical network where TSPs can be set to idle or off mode.

REFERENCES


