New light particles are commonplace in Standard-Model (SM) extensions and allow to elegantly solve problems of both conceptual and observational nature, see e.g. [1]. Depending on the couplings structure, these particles may actually be as heavy as a GeV. Remarkably, new scalars in the MeV-GeV range with larger than weak couplings to SM matter are fully compatible with the body of knowledge we have on stable matter [2], because most constraints, notably astrophysical data, apply to interactions with 1st-generation matter only. Theoretically, there is no compelling reason why these particles’ couplings should be universal across generations, or flavour-diagonal [3]. Meson or τ decays at colliders are especially suited to test such couplings, not only by definition, but also because of the large statistics and accuracy now attainable.

A common hypothesis that allows for minimal model dependence is that these light particles, once produced in the decay, escape detection. The strongest limits are obtained in missing-energy searches where the new particle is either produced in the beam interaction with a fixed target, or is the product of collisions whose initial state is very well known – e.g. e⁺e⁻ beams. For a recent comprehensive review see Ref. [2].

A prototype example of search under the above hypothesis is \( \tau \to \ell (= e, \mu) \) plus an axion-like particle (ALP, denoted by \( \phi \)) [4], performed at Mark III [5] and ARGUS [6], and on-going at Belle II [7]. At these facilities, the parent \( \tau \)'s are pair-produced at well-defined energies and their decay products are collected over a large angular acceptance, which allows for an accurate estimate of the total missing energy of the system. The dominant background to this kind of search is represented by SM processes also containing undetected particles, notably neutrinos. Hence, for the rare signal and the overwhelming background alike, the separate momenta of the pair-produced \( \tau \)'s are unknown. To pinpoint the signal, the reference strategy has historically been to estimate the signal-\( \tau \) momentum using the visible momenta on the tag side. If for the latter one assumes \( \tau \to 3\pi \nu \) (note that the 3π allow for
a high-quality vertex), the signal-τ momentum may then be estimated via the relations $\mathbf{p}_\tau \approx -\sum_{\text{tag}} \mathbf{p}_{\tau\text{tag}}$, $E_\tau \approx \sqrt{s}/2$ [6].\footnote{Here and henceforth, a hat denotes a unit vector.}

A generalisation of this technique [9], representing the current state-of-the-art, takes advantage of the ‘thrust axis’ [10,11] of the event, identified from the maximum of the ‘thrust scalar’ $T = \sum p_i - \mathbf{n}\sum |p_i|$, where $\mathbf{p}_i$ denotes all the visible momenta in the decay. This direction can be used for approximating the signal-τ momentum [7] as

$$p_\tau = \frac{\sqrt{s}}{2} (1, \mathbf{n}, \sqrt{1 - 4m_\tau^2/s}).$$  \hspace{1cm} (1)

In either case, the spectrum of the signal-side daughter lepton accompanying the φ particle is then calculated through a boost to the rest frame of the parent τ.

We propose a different approach based on the following observations: (i) signal and background decays have a common topology, consisting of visible final states plus either the elusive φ (signal), or neutrinos (backgrounds). There exists an arsenal of kinematic variables that are designed precisely for such pairwise decay topology, in particular the ‘transverse mass’ $M_{T2}$ [12,13] and Lorentz-invariant generalisations thereof [14], here collectively denoted as $M_2$; (ii) $M_2$ is a minimisation procedure in the three unknowns constituting the invisible 3-momentum on one of the two decay branches.\footnote{At lepton colliders, barring initial-state radiation, the invisible 3-momentum on the other branch is known, because the total missing momentum is.} The minimum, henceforth referred to as $M_2$-Assisted On-Shell, or ‘MAOS’, invisible momentum [15,16], is distributed around the corresponding true invisible 3-momentum. $M_2$ thereby offers a ‘built-in’ estimator of the invisible 3-momenta, separately for the two branches – which addresses the underlying challenge of the search; (iii) We expect $M_2$, as well as MAOS momenta, to show negligible correlation with variables built out of visible momenta only. Because of this small correlation all of these variables can be profitably combined. Our results show that the performance of such combination is substantially higher than the case where the different variables discussed are used individually.

We introduce our idea in the context of $\tau \rightarrow \ell + \phi$,\footnote{In all formulae it is understood that $\ell = e, \mu$. All numerics will assume $\ell = e$ for consistency with the analysis in [7].} with φ an ALP. However, the kinematic methods discussed may be applied to any other beyond-SM scenario with the same topology, e.g. lepton flavour violating couplings mediated by a spin-0 or spin-1 particle. We will make comments notably on the case of a ‘hidden photon’.

The signal of interest is $e^+e^- \rightarrow \tau (\rightarrow \ell \phi) \tau (\rightarrow 3\pi \nu)$, \hspace{1cm} (2)
where we have omitted charge specifications on the r.h.s., and we assume that the φ escapes undetected. The τ decay to three charged pions is used as a tag. We denote this channel as ‘$1 \times 3$’. The dominant, irreducible background is

$$e^+e^- \rightarrow \tau (\rightarrow \ell \nu \bar{\nu}) \tau (\rightarrow 3\pi \nu),$$  \hspace{1cm} (3)
while other channels such as $\tau (\rightarrow \pi \nu) \tau (\rightarrow 3\nu \pi)$ are all suppressed by either PID requirements, kinematic selections, event-shape analysis or track vertexing. The tag decay in eq. (2) is chosen for comparison with current state-of-the-art measurements; however, MAOS momenta can also be calculated for 3-prong tags such as $\tau (\rightarrow \ell \nu \bar{\nu})$, for which the ARGUS/thrust method is not available. We will thus also consider this ‘$1 \times 1$’ channel, whose irreducible background is $\tau (\rightarrow \ell \nu \bar{\nu})\tau (\rightarrow \ell \nu \bar{\nu})$.

Clearly, in terms of the decay topology the signal and backgrounds thus differ only by the number of invisible particles. As mentioned, this topology lends itself to the use of $M_{T2}$ and its generalizations. Such variables have been extensively applied to high-pT searches, notably of pair-produced supersymmetric particles. Instead, to our knowledge, this approach has never been considered for decays of pair-produced mesons or leptons, of the kind of interest here.

The $M_{T2}$ variable is the two-decay-chain generalization of the $M_T$ variable [17,18], used since CERN’s UA1 experiment to measure the $W$ mass in $W \rightarrow \ell \nu$, and defined from the inequality $m_W^2 \geq m_T^2 = m_1^2 + m_2^2 + 2(E_1^{\ell}E_2^{\nu} - \mathbf{p}_1^\nu \mathbf{p}_2^\nu) \equiv M_T^2$ implying that the $M_T$ endpoint allows to measure $m_W$. If one has two parents decaying to visible products with collective momenta $p_{1,2}$ for the two branches, plus invisible final states on either branch, one may generalise the above argument with $\max\{M_T(\text{branch}_1), M_T(\text{branch}_2)\}$ – the largest $M_T$ will furnish the best lower bound. While the two invisible momenta, $k_{1,2}$, are not known individually, they fulfill the constraint $k_{1T} + k_{2T} = p_T^{\text{miss}}$, where the r.h.s. denotes the measured transverse momentum imbalance. Hence, the ‘most conservative max’ one can take is the minimum over the configurations fulfilling such constraints, i.e.

$$M_{T2} = \min_{k_{1T}, k_{2T}} \left[ \max \{M_T(p_{1T}, k_{1T}), M_T(p_{2T}, k_{2T})\} \right]$$
subject to $k_{1T} + k_{2T} = p_T^{\text{miss}}$, \hspace{1cm} (4)
which defines $M_{T2}$ [12,13].

The subscript “T” in the above discussion denotes the projection onto the plane transverse to the beam direction. Such projection is unnecessary at lepton colliders like Belle II, where the full, transverse as well as longitudinal, momentum imbalance is reconstructible. We can then employ the fully Lorentz-invariant extension of $M_{T2}$, known as $M_2$ [14] (see also [19,20]). As elucidated in Ref. [20], several variations of the $M_2$ variable can actually be used for one and the same topology, depending on the kinematic constraints – i.e. on-shell mass relations – that are imposed in the minimisation and those that are not. This feature makes $M_2$ an extremely versatile tool. Since the centre-of-mass system (CMS) energy is fixed, we focus on the following definition

$$M_2 = \min_{k_1, k_2} \left[ \max \{M(p_1, k_1), M(p_2, k_2)\} \right]$$
subject to $k_1 + k_2 = p_T^{\text{miss}}$, \hspace{1cm} (5)
where s is the squared collision energy, and $p_{1,2}$ denote here and henceforth the visible- (invisible-) system total momenta on the decay branch $i = 1, 2$. One may further subject eq. (5) to the decaying parents’ on-shell mass relations, i.e. $(p_1 + k_1)^2 = (p_2 + k_2)^2 = m_\tau^2$. However, the additional constraints lead to the same minimum as our $M_2$ definition [21].\footnote{In the definition (5), $M_2$ is similar to $M_{2\text{comb}}$ devised to study $H \rightarrow \tau^+ \tau^-$ at hadron colliders [21,22], except that the constraint on the missing longitudinal momentum is not adopted.}

Similarly as for $M_{T2}$, also the $M_2$ endpoint is the parent-particle mass. Compared to $M_{T2}$, $M_2$ distributions peak at a higher value and are more populated toward the endpoint. Interestingly, one common feature is that the smaller the number of invisible particles, the more the distributions are populated toward their upper edge (for $M_{T2}$, see related discussions in [23,24]). This is displayed in Fig. 1 for the case $m_\phi = 1$ TeV.\footnote{All of Figs. 1 through 3 illustrate for a specific mass case, $m_\phi = 1$ MeV, the variables discussed in the text that rely on the use of MAOS momenta. For the sake of comparison, analogous figures for $m_\phi = 1$ GeV are collected in Fig. 6 of the Appendix.} Hence a shape analysis of
both \( M_{T2} \) and \( M_{T3} \) could in principle be used to extract information about the number of invisible particles in the event, i.e. whether the event is more signal- or background-like. Given the correlation between \( M_{T2} \) and \( M_{T3} \), for definiteness we only consider the latter in the rest of our discussion.

Importantly, the MAOS solution to the constrained minimisation in eq. (5) can be used as an estimator of the true values of \( k_{1,2} \), to be denoted as \( k_{1,2}^{\text{mas}} \) [15,16]. Similarly as in the \( M_{T2} \) case, the \( M_{2} \)-based MAOS momenta [20,25] are distributed symmetrically around the true momenta, and peak at the respective true values. A few remarks are in order. First, the \( M_{2} \)-based MAOS method turns out to be more efficient than the traditional MAOS method from \( M_{T2} \) [25]. One reason is the fact that the \( M_{T2} \)-based MAOS solution comes with a twofold ambiguity in the longitudinal components of the invisible momenta, whereas the \( M_{2} \)-based MAOS solution is unique, as all momentum components are treated on a similar footing. Second, according to the definition of the full invariant masses \( M \) in eq. (5), one further piece of information would be required: \( k_{1,2}^{2} = 0 \). We set \( k_{1,2}^{2} = 0 \), which usefully ensures the inequality \( M \leq m_{T2} \) for both signal and background events [15,16]. At face value, \( k_{1,2}^{2} = 0 \) is a bad guess, because e.g. for the background channel \( \ell \nu \nu \) one has \( k_{1,2}^{2} = m_{\nu \nu}^{2} \), which peaks around 1 GeV. We inspected more refined ansätze, including the unfeasible case where one uses the truth-level \( k_{1,2}^{2} \) values. We found that the \( M_{2} \) distributions do not depend significantly on this choice, in agreement with existing literature [15,16], although this issue may warrant further scrutiny.

With the thus-defined \( K_{1,2}^{\text{mas}} \) we can construct additional variables that would require knowledge of the invisible momenta, and that offer criteria for signal-background discrimination. A first example is \( R_{p}^{\text{mas}} \) in the \( 1 \times 3 \) channel. In the \( \tau \) rest frame (RF), obtained through either the MAOS or thrust methods, this quantity is equivalent to the signal-side invisible momentum. In Fig. 2 we thus compare the distribution for \( R_{p}^{\text{mas}} \) with the corresponding quantity obtained through the thrust, for \( m_{A} = 1 \) MeV, representative of the small-mass case. For comparison, the case \( m_{A} = 1 \) GeV is shown in Fig. 6 in the Appendix. In either case we restrict to the \( 1 \times 3 \) channel. MAOS and thrust achieve a comparable separation between the signal and the background distributions for \( m_{A} = 1 \) GeV, whereas MAOS performs better for small \( m_{A} \), as quantified by the full analysis, to be discussed at the end of the paper.

One further quantity constructible from \( K_{1,2}^{\text{mas}} \) is the ratio

\[
\xi_{k} = \frac{\min[|k_{1}|, |k_{2}|]}{\max[|k_{1}|, |k_{2}|]} \in [0, 1], \tag{6}
\]

with \( \xi_{k} \) denoting the corresponding ratio calculated with \( K_{1,2}^{\text{mas}} \). This variable is reminiscent of the \( R_{p} \) ratio pointed out in Ref. [23], with two differences. First, \( R_{p} \) is constructed in terms of the visible momenta, whereas \( \xi_{k} \) requires the invisible ones, separately for the two branches – which is precisely what MAOS provides. Second, \( R_{p} \) is a ‘max-over-min’ ratio, implying the non-compact domain [1, \( \infty \)], i.e. a long distribution tail. The difference between signal and background is thus ‘diluted’ over this tail. Conversely, \( \xi_{k} \) spans the compact domain [0, 1], which enhances the shape difference between signal and background. The \( \xi_{k} \) distribution performs best of all and is shown in Fig. 3 (first panel) for the case \( m_{A} = 1 \) MeV. Note that \( \xi_{k,0} \) could be defined in the lab or CMS frames. We used the CMS-frame definition, where the slope differences are more pronounced. For purposes of comparison, the \( \xi_{k} \), \( R_{k} \), and \( R_{p} \) distributions are shown in the remaining panels of Fig. 3. Besides, all of \( \xi_{k,0}, R_{k,0}, R_{p,0} \) are also shown in Fig. 6 of the Appendix for the case \( m_{A} = 1 \) GeV. The underlying rationale of eq. (6) is that this ratio is expected to be closer to unity for the \( 1 \times 1 \)-channel background (4th entry in the legend), because of the symmetric decay chains. By the same argument, we also expect the distribution ‘slope’ to decrease roughly with the number of invisibles. The \( \xi_{k} \) histograms, and to a lesser extent the \( \xi_{p} \) ones, display both of these features.

We shortly discuss other known variables that do not require MAOS momenta, and that show a small enough correlation with those discussed so far. One popular example for lepton colliders is the recoil mass [26,27], defined as \( M_{\text{rec}}^{2} = \left(P_{\text{CMS}}^{2} - P_{1}^{2} - P_{2}^{2}\right)^{2} \) i.e.

![Fig. 1. Comparison between \( M_{T2} \) and \( M_{T3} \) in the separation of signals and backgrounds (full vs. dashed lines) for the \( 1 \times 3 \) (blue) and the \( 1 \times 1 \) (orange) topologies. We show the case \( m_{A} = 1 \) MeV.](image1)

![Fig. 2. Distribution for \( p_{T}^{\text{mas}} := |p_{T}^{\text{mas}}| \) and for the corresponding quantity obtained through the thrust, for the \( 1 \times 3 \) channel and \( m_{A} = 1 \) MeV.](image2)

![Fig. 3. Comparison between the ratio variables \( \xi_{k,0}, R_{k,0}, R_{p,0} \) discussed below eq. (6). Line conventions as in Fig. 1. We show the case \( m_{A} = 1 \) MeV.](image3)
the invariant mass of the full invisible system. Since there are more invisible particles in the background, they ‘typically’ have a larger invariant mass in the signal case. This property is clearly visible in the $M_{\text{recoll}}$ distribution in Fig. 4 (first panel), for both the $1 \times 3$ and the $1 \times 1$ cases. One may reverse the argument for the variable $E_{\text{miss}} \equiv |P_{\text{miss}}|$: in the $\tau$-pair rest frame, the invisible particles would be boosted along the momentum direction of the parent $\tau$, so their three-momenta will partially cancel those from the other decay chain. This cancellation will be the more efficient, the more symmetric the decay. Hence one may expect a ‘thicker’ tail for the signal decay $\ell^+ \ell^- + \nu\bar{\nu}$ than for the corresponding background $\ell^+\ell^- + \nu\bar{\nu}$. In practice, the $E_{\text{miss}}$ discriminating power is inferior to $M_{\text{recoll}}$’s, as shown by the second panel of Fig. 4.\footnote{Other examples of variables relying on the visible kinematics only were investigated in [28]. We do not include them in our analysis, as we do not expect them to modify appreciably our conclusions, as is also the case for $M_{\text{recoll}}$ and $E_{\text{miss}}$.}

Before presenting our main analysis, we collect details about our setup. We generate $e^+e^- \rightarrow \tau^+\tau^-$ using MadGraph. Tag-side decays and backgrounds are obtained through TauDecay [29], whereas signal-side decays are populated as phase space through ROOT. We generate about $1.5 \times 10^5$ events per process. In order to populate the phase space in a similar way as Ref. [7], we apply the cut $0.8 \leq T \leq 0.99$ on the thrust scalar. We also inspected the effect of including further cuts on the total visible energy and on the invariant mass of the 3-prong system, used in the Belle-II analysis [7] for the suppression of reducible backgrounds, and found it to be negligible in our case. Momentum smearing due to detector effects is typically $\lesssim 1\%$ and we safely neglect this effect. To be more exact, event distributions are vastly more populated for $p_T > 0.2$ GeV/c [30], which is the region where the below-1\% momentum-smearing figure holds. Our numerical analysis uses the public library YAM2 [31] and the TMVA [32] class available in ROOT. We restrict to phase-space decays for comparison with Ref. [7], and also for the following reason. The variables discussed above are insensitive to the angular distribution of the new particle. The cuts implemented to mimic the search in Ref. [7] will not modify angular distributions either, as these cuts affect invisible masses or momentum magnitudes. We also note that, since our decay of interest is 2-body, the decay amplitude is isotropic and so is the differential decay rate [33]. As a consequence, our results apply equally to the case of an ALP and of a hidden photon, with coupling chirality whatever. A separate, interesting question would be to construct “invisible-spin-savvy” variables, which exploit MAOS or thrust momenta in spin-sensitive kinematic variables (as in e.g. [34]), in order to tell apart different coupling assumptions.

We next discuss the main analysis. We first note that $M_{\text{recoll}}, \mathbf{E}, p_T, M_{\text{recoll}}$ and $E_{\text{miss}}$ can be unambiguously calculated for the $1 \times 3$ and $1 \times 1$ cases alike,\footnote{On the other hand, application of $p_{\text{miss}}$ and $p_T$ is not straightforward in the $1 \times 1$ case, because the symmetry of the topology implies a combinatorial ambiguity, introducing a separate source of uncertainty. Including these variables in this case requires a dedicated study (see e.g. discussion in [35]).} and their distributions depend, to different degrees, on the number of invisibles in the decay. We thus collectively denote this ensemble as ‘invisible-savvy’ variables, and construct a classifier that we refer to as $\mathcal{I}_{\text{Sy}}$. Note that $\mathcal{I}_{\text{Sy}}$ does not include $p_T$. We then consider the following cases: $(a)$ $p_T$ alone, on $1 \times 3$ decays; $(b)$ $p_{\text{miss}}$ alone, on $1 \times 3$ decays; $(c)$ $\mathcal{I}_{\text{Sy}}$ alone, on $1 \times 3$ decays; $(abc)$ $p_T$ + $p_{\text{miss}}$ + $\mathcal{I}_{\text{Sy}}$ combined, on $1 \times 3$; $(d)$ $\mathcal{I}_{\text{Sy}}$, on $1 \times 1$. With case $(a)$ we reproduce the results in Ref. [7] and thus validate our setup; the comparisons $(a)$ vs. $(b)$ vs. $(c)$ vs. $(abc)$ show the improvement achievable on the $1 \times 3$ channel alone with the different variables, and with their combination; $(c)$ vs. $(d)$ shows the relative performance of the two considered channels. In this last comparison, we only consider the $\mathcal{I}_{\text{Sy}}$ classifier, keeping in mind footnote 8.

We refrain from including an $(abcd)$ case, which would show the ‘maximal’ improvement achievable from both the use of two channels in lieu of one, and the use of more variables. In fact, a reliable combination of $1 \times 3$ and $1 \times 1$ is not straightforward, because of the different tag, and should be performed on actual data.

Fig. 5 (first panel) presents a comparison of the performance profiles for the cases discussed above, i.e. $(a)$ to $(d)$, in the plane of signal efficiency vs. background rejection, yielding the ‘ROC’ curve. In the single-variable cases $(a)$ and $(b)$, we use a cut-based approach rather than a BDT/NN, although the optimal cut is obtained by TMVA. We see that, for $m_{\ell} = 1$ MeV, $p_{\text{miss}}$ has a larger area-under-curve (AUC) than $p_T$ for any signal efficiency. For $m_{\ell} = 1$ GeV, the two AUCs are comparable, and we show the corresponding plot in Fig. 7 (left) in the Appendix. In the remaining panels of this figure we also show, for the sake of possible reproducibility, the BDT response for both cases $m_{\ell} = 1$ MeV and 1 GeV.

Our $S/B$ separation can be translated into an estimate of the upper limit (UL) on $B(\tau \rightarrow e\phi)$ achievable with a given Belle-II luminosity. To determine such limit we proceed as follows. Given a sample of $N_I$ signal and $N_B$ background events, we weight the sample background can be determined as $w_B = B/N_B$ where, in our case, $B = \sigma_{\tau\tau} \times B(\tau \rightarrow bkg) \times \mathcal{L}$, with $\sigma_{\tau\tau}$ the total $\tau^+\tau^-$ production cross-section and $\mathcal{L}$ the luminosity. We need the weight $w_i$ for the signal sample, from which we may estimate the statistical significance as $\sigma_{\mathcal{W}_i} \sim S/\sqrt{S+B} = w_i \sqrt{N_s w_s + N_B w_B}$. The $\mathcal{S}_{\mathcal{W}_i}$ value corresponding to a 95% confidence-level (CL) exclusion is given by $\sigma_{\mathcal{W}_i} = 1.96$ [33], that we invert in terms of $w_i$. With fixed $N_B$, one may proceed iteratively starting from $w_i^{(0)} = w_s$ and decreasing $w_i^{(l)}$ till the desired equality is satisfied.) From $S = w_i N_s$, one finally obtains the corresponding signal branching-ratio value through $B(\tau \rightarrow e\phi)/B(\tau \rightarrow bkg) = S/B$.

This procedure is independent from the template-fit method used for the on-going Belle-II analysis [7], hence the agreement with [7] of the $p_T$-based upper-limit curve shown in Fig. 5. This is a non-trivial check of our approach. We also verified that, after the classifier cut, $B$ is large enough that the above approximate relations are valid [36].

Before concluding with the main analysis results we note that our approach has a wide range of applicability – be it to searches of new decays to invisibles or to improving the knowledge of background decays to invisibles – to the extent that the ‘pairwise-decay’ topology is the same. Besides, if one restricts to transverse variables, a similar approach may also be applied for meson or $\tau$ decays at hadron colliders.

We conclude by presenting the 95% CL upper limit on the signal branching ratio (BR) as a function of the new light particle mass, for different assumed Belle-II luminosities $\mathcal{L}$, and with the different classifiers discussed. These upper limits are summarised

---

Fig. 4. The variables $M_{\text{recoll}}$ and $E_{\text{miss}}$. Line conventions as in Fig. 1. We show the case $m_{\ell} = 1$ MeV.
in the middle and rightmost panels of Fig. 5. In particular, the former shows, for the integrated luminosity $\mathcal{L} = 0.1 \text{ ab}^{-1}$, a comparison among the cases (a), (b), (c), (d), (abc) discussed above; conversely, the last panel of Fig. 5 focuses on the evolution of the expected upper limit with luminosity, and shows the cases $\mathcal{L} = [0.1, 1.50] \text{ ab}^{-1}$, corresponding to the dataset accumulated as of Summer 2021, the dataset anticipated before the 2022 shutdown, and the target Belle-II dataset, respectively. As also suggested by the ROC curves, $p_{e\text{mass}}$ allows for a better BR limit than $p_{e\text{brust}}^\text{mass}$ for small $m_\phi \lesssim 0.1 \text{ GeV}$. For example, this improvement is a factor of $\approx 1.8$ for $m_\phi \approx 0$ and $\mathcal{L} = 0.1 \text{ ab}^{-1}$. In short, for an ALP or hidden vector of small $m_\phi \lesssim 1 \text{ MeV}$, we anticipate that application of our full strategy to the $1 \times 3$ channel alone will lead to a 95%-CL limit of around

$$B(\tau \rightarrow e^+ \phi) \lesssim [5.4 \cdot 10^{-5}, 1.7 \cdot 10^{-5}, 2.4 \cdot 10^{-6}],$$

for $\mathcal{L} = [0.1, 1.50] \text{ ab}^{-1}$, (7)

to be compared with $B(\tau \rightarrow e^+ \phi) \lesssim [1.3 \cdot 10^{-4}, 4.0 \cdot 10^{-5}, 5.7 \cdot 10^{-6}]$ with the thrust method alone. As a consequence, our strategy improves by a factor close to 3 the limit achievable with the strategy currently in place within Belle II. We thus expect a Belle-II limit on $\text{BR}(\tau \rightarrow e^+ \text{invisible})$ stronger than the existing ARGUS limit [6] by a factor of respectively 50, 170, 1150 with $0.1, 1, 50 \text{ ab}^{-1}$ Belle-II data.

**Declaration of competing interest**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**Acknowledgements**

DG warmly acknowledges Diego Redigolo for several related discussions following his talk [37], and Justine Serrano for comments. Exchanges with Aleks Smolkovic and Jure Zupan are also acknowledged. This project has received funding from the ANR under contract n. 202650 (PRC ‘GammaRare’), IBS under the project code, IBS-R018-D1 and from the European Union’s Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 101026516.

**Appendix A. Additional figures**

Fig. 6. Same as Figs. 1 + 2 + 3 (with the corresponding panels in the same order), but for $m_\phi = 1 \text{ GeV}$.

---

**Fig. 5.** Left: ROC curves for the different classifiers discussed in the text, in the case $m_\phi = 1 \text{ MeV}$. Right: 95% CL upper limits on the signal branching ratios in units of the $\tau \rightarrow \ell \nu \nu$ one. We consider the different classifiers discussed in the text (see legend) and for each of them three reference luminosities (see plot headers).
Fig. 7. First panel: ROC curve for the case $m_\phi = 1$ GeV, to be compared with the 1-MeV case, shown in the first panel of Fig. 5 in the main text. Second and third panels: distributions in the BDT response, for both cases $m_\phi = 1$ MeV and 1 GeV.

References

[1] J. Beacham, et al., Physics beyond colliders at CERN: beyond the standard model working group report, J. Phys. G 47 (2020) 010501, arXiv:1901.09966 [hep-ph], 2020.
[2] G. Lanfranchi, M. Pospelov, P. Schuster, The search for freely-interacting particles, https://doi.org/10.1146/annurev-nucl-102419-055056, arXiv:2011.02157 [hep-ph], 2020.
[3] H. Georgi, D.B. Kaplan, L. Randall, Manifesting the invisible axion at low-energies, Phys. Lett. B 169 (1986) 73.
[4] L. Calibbi, D. Redigolo, R. Ziegler, J. Zupan, Looking forward to lepton-flavor-violating ALPs, arXiv:2006.04795 [hep-ph], 2020.
[5] R.M. Baltrusaitis, et al., MARK-III, $t$ lepton branching ratios and a search for Coldstone decay, Phys. Rev. Lett. 55 (1985) 1842.
[6] H. Albrecht, et al., ARGUS, A search for lepton flavor violating decays tau $\rightarrow e$ alpha, tau $\rightarrow \mu$ alpha, Z. Phys. C 68 (1995) 25.
[7] F. Tenenuchi, M. Garcia-Hernandez, T. Kraetzschmar, P.K. Rados, E. De La Cruz-Burelo, A. De Yta-Hernandez, I. Heredia de la Cruz, A. Rostomyan, First results and prospects for tau LPV decay $\tau \rightarrow e\nu\nu$(invisible) at Belle II, in: Proceedings, 40th International Conference on High Energy Physics (ICHEP2020), Prague, Czechia, July 28 - August 6, 2020, PoS ICHEP2020 (2020) 288.
[8] D.A. Bryman, S. Ito, S. Shrock, Upper limits on branching ratios of decays $\tau \rightarrow \nu\gamma\gamma$ and $\tau \rightarrow eX$, arXiv:2106.02451 [hep-ph], 2021.
[9] F. Abudirhem, et al., Belle-II, $t$ lepton mass measurement at Belle II, arXiv:2008.04665 [hep-ex], 2020.
[10] S. Brandt, C. Peyrou, R. Sosnowski, A. Wrobleski, The principal axis of jets. An attempt to analyze high-energy collisions as two-body processes, Phys. Lett. 12 (1964) 52.
[11] E. Farhi, A QCD test for jets, Phys. Rev. Lett. 39 (1977) 1587.
[12] C.G. Lester, D.J. Summers, Measuring masses of semivisibly decaying particles pair produced at hadron colliders, Phys. Lett. B 463 (1999) 99, arXiv:hep-ph/9906349 [hep-ph].
[13] A. Barr, C. Lester, P. Stephens, m(T2): the truth behind the glamour, J. Phys. G 29 (2003) 2343, arXiv:hep-ph/0304226 [hep-ph].
[14] A.J. Barr, T.J. Kho, P. Konar, K. Kong, C.G. Lester, K.T. Matchev, M. Park, Guide to transverse projections and mass-constraining variables, Phys. Rev. D 84 (2011) 095031, arXiv:1105.2977 [hep-ph].
[15] W.S. Cho, K. Choi, Y.G. Kim, C.B. Park, M(T2)-assisted on-shell reconstruction of missing momenta and its application to spin measurement at the LHC, Phys. Rev. D 79 (2009) 031701, arXiv:0810.4853 [hep-ph].
[16] C.B. Park, Reconstructing the heavy resonance at hadron colliders, Phys. Rev. D 84 (2011) 096001, arXiv:1106.6087 [hep-ph].
[17] J. Smith, W.L. van Neerven, J.A.M. Vermaseren, The transverse mass and width of the $W$ boson, Phys. Rev. Lett. 50 (1983) 1738.
[18] V.D. Barger, A.D. Martin, R.J.N. Phillips, Perpendicular $\nu_\mu$ mass from $W$ decay, in: 11th International Symposium on Lepton and Photon Interactions at High Energies (Tbilisi, New York, August 4-9, 1983, Z. Phys. C 21 (1983) 99.
[19] G.G. Ross, M. Serna, Mass determination of new states at hadron colliders, Phys. Lett. B 665 (2008) 212, arXiv:0712.0943 [hep-ph].
[20] W.S. Cho, J.S. Gainer, D. Kim, K.T. Matchev, F. Moorrtgat, L. Pape, M. Park, On-shell constrained $M_2$ variables with applications to mass measurements and topology disambiguation, J. High Energy Phys. 08 (2014) 070, arXiv:1401.1449 [hep-ph].
[21] P. Konar, A.K. Swain, Mass reconstruction with $M_2$ under constraint in semi-invisible production at a hadron collider, Phys. Rev. D 93 (2016) 015021, arXiv:1509.00528 [hep-ph].
[22] P. Konar, A.K. Swain, Reconstructing semi-invisible events in resonant tau pair production from Higgs, Phys. Lett. B 757 (2016) 211, arXiv:1602.00552 [hep-ph].
[23] K. Agashe, D. Kim, D.G.E. Walker, L. Zhu, Using $M_2$ to distinguish dark matter stabilization symmetries, Phys. Rev. D 84 (2011) 055020, arXiv:1012.4640 [hep-ex], 2012.
[24] G.F. Guidice, B. Gripaios, R. Mahbubani, Counting dark matter particles in LHC events, Phys. Rev. D 85 (2012) 075019, arXiv:1108.1800 [hep-ph].
[25] D. Kim, K.T. Matchev, F. Moorrtgat, L. Pape, Testing invisible momentum ansatze in missing energy events at the LHC, J. High Energy Phys. 08 (2017) 102, arXiv:1703.06887 [hep-ph].
[26] H. Li, K. Ito, R. Posch, F. Richard, M. Ruan, Y. Takubo, H. Yamamoto, ILD Design Study Group, HZ recoil mass and cross section analysis in ILD, arXiv:1202.1439 [hep-ex], 2012.
[27] K. Fuji, et al., Physics case for the international linear collider, arXiv:1506.05992 [hep-ex], 2015.
[28] E. De La Cruz-Burelo, A. De Yta-Hernandez, M. Hernandez-Villanueva, New method for beyond the Standard Model invisible particle searches in tau lepton decays, Phys. Rev. D 102 (2020) 115001, arXiv:2007.08239 [hep-ph].
[29] K. Hagiwara, T. Li, K. Kawatari, J. Nakamura, TauDecay: a library to simulate polarized tau decays via FeynRules and MadGraph5, Eur. Phys. J. C 73 (2013) 2489, arXiv:1212.5247 [hep-ph].
[30] V. Bertacchi, et al., Belle II Tracking Group, Track finding at Belle II, Comput. Phys. Commun. 259 (2021) 107610, arXiv:2003.12466 [physics.ins-det].
[31] C.B. Park, YAM2: yet another library for the $M_2$ variables using sequential quadratic programming, Comput. Phys. Commun. 264 (2021) 107967, arXiv:2007.15537 [hep-ph].
[32] A. Hooker, et al., TMVA - toolkit for multivariate data analysis, arXiv:physics/0703039, 2007.
[33] P.A. Zyla, et al., Particle Data Group, Review of particle physics, PTEP 2020 (2020) 083C01.
[34] D. Guadagnoli, C.B. Park, $M_2$-reconstructed invisible mass as spin analyzers, and an application to top polarization, J. High Energy Phys. 01 (2014) 030, arXiv:1308.2226 [hep-ph].
[35] K. Choi, D. Guadagnoli, C.B. Park, Reducing combinatorial uncertainties: a new technique based on MTZ variables, J. High Energy Phys. 11 (2017) 117, arXiv:1109.2201 [hep-ph].
[36] P.N. Bhattachrullu, S.P. Martin, J.D. Wells, Criteria for projected discovery and exclusion sensitivities of counting experiments, Eur. Phys. J. C 81 (2021) 123, arXiv:2009.07249 [physics.data-an].
[37] D. Redigolo, Linking axions/ALPs with flavour, and searches at flavour experiments, in: GDR Intensity Frontier Annual Meeting, 2020, https://indico.in2p3.fr/event/22193.